

On Asymptotic Optimality in Sequential Changepoint Detection: Non-iid Case

Alexander G. Tartakovsky

Somers, CT, USA.

e-mail: alexg.tartakovsky@gmail.com

Abstract: We consider a sequential Bayesian changepoint detection problem for a general stochastic model, assuming that the observed data may be dependent and non-identically distributed and the prior distribution of the change point is arbitrary, not necessarily geometric. Tartakovsky and Veeravalli (2004) developed a general asymptotic theory of changepoint detection in the non-iid case and discrete time, and Baron and Tartakovsky (2006) in continuous time assuming the certain stability of the log-likelihood ratio process. This stability property was formulated in terms of the r -quick convergence of the normalized log-likelihood ratio process to a positive and finite number, which can be interpreted as the limiting Kullback–Leibler information between the “change” and “no change” hypotheses. In these papers, it was conjectured that the r -quick convergence can be relaxed in the r -complete convergence, which is typically much easier to verify in particular examples. In the present paper, we justify this conjecture by showing that the Shiryaev change detection procedure is nearly optimal, minimizing asymptotically (as the probability of false alarm vanishes) the moments of the delay to detection up to order r whenever r -complete convergence holds. We also study asymptotic properties of the Shiryaev–Roberts detection procedure in the Bayesian context.

Keywords and phrases: Asymptotic Optimality; Changepoint Problems; Complete Convergence; Hidden Markov Models; Markov Process; r -quick Convergence.

Received January 2016.

1. Introduction

In the beginning of the 1960s, [Shiryaev \(1961, 1963\)](#) developed a Bayesian sequential changepoint detection (quickest disorder detection) theory in the iid case assuming that the observations are independent and identically distributed (iid) according to a distribution F pre-change and another distribution G post-change and with the prior distribution of the change point being geometric. In particular, [Shiryaev \(1963\)](#) proved that the detection procedure that is based on thresholding the posterior probability of the change being active before the current time is strictly optimal, minimizing the average delay to detection in the class of procedures with a given probability of false alarm. [Tartakovsky and Veeravalli \(2005\)](#) generalized Shiryaev’s theory for the non-iid case that covers very general discrete-time non-iid stochastic models and a wide class of prior distributions that include distributions with both exponential tails and heavy tails. In particular, it was proved that the Shiryaev detection procedure is asymptotically optimal – it minimizes the average delay to detection as well as higher moments of the detection delay as the probability of a false alarm vanishes. [Baron and Tartakovsky \(2006\)](#) developed an asymptotic Bayesian theory for general continuous-time stochastic processes.

The key assumption in general asymptotic theories developed in [Baron and Tartakovsky \(2006\)](#); [Tartakovsky and Veeravalli \(2005\)](#) is a certain stability property of the log-likelihood ratio process between the “change” and “no-change” hypotheses, which was expressed in the form of the strong law of large numbers with a positive and finite number and its strengthened r -quick version. However, it is not easy (and in fact can be quite difficult) to verify r -quick convergence in particular applications and examples. For this reason, it was conjectured in [Baron and Tartakovsky \(2006\)](#); [Tartakovsky and Veeravalli \(2005\)](#) that essentially the same asymptotic results may be obtained under a weaker r -complete version of the strong law of large

numbers for the log-likelihood ratio. In fact, in most examples provided in [Baron and Tartakovsky \(2006\)](#); [Tartakovsky and Veeravalli \(2005\)](#) and in the recent book by [Tartakovsky et al. \(2014\)](#), verification of the r -quick convergence is replaced by verification of the r -complete convergence. The main goal of the present article is to confirm this conjecture, proving that the Shiryaev changepoint detection procedure is asymptotically optimal under the r -complete convergence condition for the suitably normalized log-likelihood ratio process.

The rest of the article is organized as follows. We formulate a general Bayesian changepoint detection problem and present some preliminary results in Section 2. In Section 3, we consider the Shiryaev change detection procedure in detail and prove that it is asymptotically optimal under mild conditions associated with the r -complete convergence of the properly normalized log-likelihood ratio. In Section 4, we discuss asymptotic properties and derive operating characteristics of another popular change detection procedure, the Shiryaev–Roberts procedure, and show that in general it is not asymptotically optimal in the Bayesian context, but preserves asymptotic optimality properties under certain conditions of the prior distribution. In Section 5, we provide examples of interesting cases where the conditions that we posit in Section 3 and Section 4 are satisfied. In Section 6, we conclude the paper by discussing the relevance of our results and providing additional remarks. Most of the proofs are presented in the main body of the paper, but proofs of some lemmas are given in the Appendix.

2. Problem setup and preliminaries

In the following, we deal only with discrete time $t = n \in \mathbb{Z}_+ = \{0, 1, 2, \dots\}$. The continuous time case $t \in \mathbb{R}_+ = [0, \infty)$ is more “delicate” and will be considered elsewhere. Having said that, let $(\Omega, \mathcal{F}, \mathcal{F}_n, \mathbb{P})$, $n \in \mathbb{Z}_+$ be a filtered probability space, where the sub- σ -algebra $\mathcal{F}_n = \sigma(\mathbf{X}^n)$ of \mathcal{F} is assumed to be generated by the process $\mathbf{X}^n = \{X_t\}_{1 \leq t \leq n}$ observed up to time n . Let \mathbb{P}_0 and \mathbb{P}_∞ be two probability measures defined on this space, which are assumed to be mutually locally absolutely continuous, so that the restrictions of these measures \mathbb{P}_0^n and \mathbb{P}_∞^n to the sigma-algebras \mathcal{F}_n are mutually absolutely continuous for all $n \geq 1$.

We are interested in the following changepoint problem. In a “normal” mode, the observed process X_n follows the measure \mathbb{P}_∞ , and at an unknown time ν ($\nu \geq 0$) something happens and X_n follows the measure \mathbb{P}_0 . The goal is to detect the change as soon as possible after it occurs, subject to a constraint on the risk of false alarms. The exact optimality criteria will be specified in Section 2.2.

2.1. A general changepoint model

Let $p_j(\mathbf{X}^n)$, $j = \infty, 0$ denote densities of \mathbb{P}_j^n (with respect to some non-degenerate σ -finite measure), where $\mathbf{X}^n = (X_1, \dots, X_n)$ is the sample of size n . For a fixed $\nu \in \mathbb{Z}_+$, the change induces a probability measure \mathbb{P}_ν (correspondingly density $p_\nu(\mathbf{X}^n) = p(\mathbf{X}^n|\nu)$), which is a combination of the pre- and post-change densities:

$$p_\nu(\mathbf{X}^n) = p_\infty(\mathbf{X}^\nu) \cdot p_0(\mathbf{X}_{\nu+1}^n|\mathbf{X}^\nu) = \prod_{i=1}^{\nu} p_\infty(X_i|\mathbf{X}^{i-1}) \cdot \prod_{i=\nu+1}^n p_0(X_i|\mathbf{X}^{i-1}), \quad (2.1)$$

where $\mathbf{X}_m^n = (X_m, \dots, X_n)$ and $p_j(X_n|\mathbf{X}^{n-1})$ is the conditional density of X_n given \mathbf{X}^{n-1} . In the sequel we assume that ν is the serial number of the last pre-change observation. Note that in general the conditional densities $p_0(X_i|\mathbf{X}^{i-1})$, $i = \nu + 1, \nu + 2, \dots$ may depend on the changepoint ν , i.e., $p_0(X_i|\mathbf{X}^{i-1}) = p_0^{(\nu)}(X_i|\mathbf{X}^{i-1})$ for $i > \nu$. Certainly the densities $p_j(X_i|\mathbf{X}^{i-1}) = p_{j,i}(X_i|\mathbf{X}^{i-1})$, $j = 0, \infty$ may depend on i .

In a particular *iid case*, addressed in detail in the past the observations are independent and identically distributed (iid) with density $f_\infty(x)$ in the normal (pre-change) mode and with another density $f_0(x)$ in the abnormal (post-change) mode, i.e., in this case, (2.1) holds with $p_\infty(X_i|\mathbf{X}^{i-1}) = f_\infty(X_i)$ and $p_0(X_i|\mathbf{X}^{i-1}) = f_0(X_i)$.

We are interested in a Bayesian setting where the change point ν is assumed to be a random variable independent of the observations with prior probability distribution $\Pi_n = P(\nu \leq n)$, $n \in \mathbb{Z}_+$. We also write $\pi_k = P(\nu = k)$ for the probability on non-negative integers, $k = 0, 1, 2, \dots$. Formally, we allow the change point ν to take negative values too, but the detailed distribution for $k < 0$ is not important. The only value we need is the cumulative probability $q = P(\nu < 0)$. The probability $P(\nu \leq 0) = q + \pi_0$ is the probability of the “atom” associated with the event that the change already took place before the observations became available.

In the past, the typical choice for the prior distribution was (zero modified) geometric distribution,

$$P(\nu < 0) = q \quad \text{and} \quad P(\nu = k) = (1 - q)\rho(1 - \rho)^k \quad \text{for } k = 0, 1, 2, \dots, \quad (2.2)$$

where $0 \leq q < 1$, $0 < \rho < 1$.

In the rest of the paper, we consider an arbitrary prior distribution that belongs to the class of distributions that satisfy the following condition:

C. For some $0 \leq \mu < \infty$,

$$\lim_{n \rightarrow \infty} \frac{|\log(1 - \Pi_n)|}{n} = \mu. \quad (2.3)$$

In the case that $\mu = 0$, we assume in addition that for some $r \geq 1$

$$\sum_{k=0}^{\infty} \pi_k |\log \pi_k|^r < \infty. \quad (2.4)$$

If $\mu > 0$, then the prior distribution has an exponential right tail. Such distributions, as geometric and discrete versions of gamma and logistic distributions, i.e., models with bounded hazard rate, belong to this class. In this case, condition (2.4) holds automatically. If $\mu = 0$, then the distribution has a heavy tail, i.e., such a distribution belongs to the model with a vanishing hazard rate. However, we cannot allow this distribution to have a tail that is too heavy, which is guaranteed by condition (2.4).

2.2. Optimality criteria

Any sequential detection procedure is a stopping time T for the observed process $\{X_n\}_{n \in \mathbb{Z}_+}$, i.e., T is an extended random variable, such that the event $\{T = n\}$ belongs to the sigma-algebra \mathcal{F}_n . A false alarm is raised whenever $T \leq \nu$. A good detection procedure should guarantee a small delay to detection $T - \nu$ provided that there is no false alarm, while the rate (or risk) of false alarms should be kept at a given, usually low level.

Let P_k and E_k denote the probability and the corresponding expectation when the change occurs at time $\nu = k \in \mathbb{Z}_+$. In what follows, P^π denotes the probability measure on the Borel sigma-algebra in $\mathbb{R}^\infty \times \mathbb{N}$ defined as $P^\pi(\mathcal{A} \times J) = \sum_{k \in J} \pi_k P_k(\mathcal{A})$ for $\mathcal{A} \in \mathcal{B}(\mathbb{R}^\infty)$, $J \subseteq \mathbb{N}$ and E^π denotes the expectation with respect to P^π .

In a Bayesian setting, the risk associated with the delay to detection is usually measured by the average delay to detection

$$E^\pi(T - \nu | T > \nu) = \frac{\sum_{k=0}^{\infty} \pi_k E_k(T - k | T > k) P_\infty(T > k)}{1 - \text{PFA}(T)} \quad (2.5)$$

and the risk associated with a false alarm by the weighted probability of false alarm (PFA) defined as

$$\text{PFA}(T) = \mathbb{P}^\pi(T \leq \nu) = \sum_{k=1}^{\infty} \pi_k \mathbb{P}_\infty(T \leq k). \quad (2.6)$$

In (2.5) and (2.6) we use the fact that $\mathbb{P}_k(T > k) = \mathbb{P}_\infty(T > k)$ and $\mathbb{P}_k(T \leq k) = \mathbb{P}_\infty(T \leq k)$ for $k \in \mathbb{Z}_+$ and that $\mathbb{P}_\infty(T \leq 0) = 0$.

For $0 < \alpha < 1$, let $\mathbb{C}_\alpha = \{T : \text{PFA}(T) \leq \alpha\}$ be a class of detection procedures for which the weighted probability of false alarm does not exceed the predefined level α . In a Bayesian setting, the goal is to find an optimal procedure that minimizes in the class \mathbb{C}_α the average delay to detection, i.e.,

$$\text{find } T_{\text{opt}} \in \mathbb{C}_\alpha \text{ such that } \mathbb{E}^\pi(T_{\text{opt}} - \nu | T_{\text{opt}} > \nu) = \inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi(T - \nu | T > \nu).$$

However, except for the iid case, the solution of this problem is not tractable. For this reason, we address the asymptotic problem of minimizing the average detection delay as α approaches zero. For practical purposes, it is also interesting to consider the problem of minimizing higher moments of the detection delay $\mathbb{E}^\pi[(T - \nu)^m | T > \nu]$ for some $m \geq 1$, i.e., to find a first-order asymptotically optimal detection procedure $T_o \in \mathbb{C}_\alpha$ that satisfies

$$\lim_{\alpha \rightarrow 0} \frac{\mathbb{E}^\pi[(T_o - \nu)^m | T_o > \nu]}{\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi[(T - \nu)^m | T > \nu]} = 1. \quad (2.7)$$

2.3. Change detection procedures

Let “ $H_k : \nu = k$ ” and “ $H_\infty : \nu = \infty$ ” be the hypotheses that the change occurs at the point $0 \leq k < \infty$ and that the change never happens, respectively. Then, using (2.1), we obtain that the likelihood ratio (LR) between these hypotheses when the sample $\mathbf{X}^n = (X_1, \dots, X_n)$ is observed is

$$\frac{d\mathbb{P}_k^n}{d\mathbb{P}_\infty^n} = \prod_{i=k+1}^n \frac{p_0(X_i | \mathbf{X}^{i-1})}{p_\infty(X_i | \mathbf{X}^{i-1})}, \quad k < n.$$

Write $\mathcal{L}_i = p_0(X_i | \mathbf{X}^{i-1})/p_\infty(X_i | \mathbf{X}^{i-1})$ and introduce the normalized average (weighted) LR

$$\Lambda_n = \frac{1}{\mathbb{P}(\nu \geq n)} \left(q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i \right), \quad n \in \mathbb{Z}_+.$$

Note that $\Lambda_0 = q/(1 - q)$. Let $g_n = \mathbb{P}(\nu < n | \mathbf{X}^n)$ stand for the posterior probability of the change being in effect up to time n . Shiryaev (1963) proved that, in the iid case, the detection procedure $T_a = \inf \{n : g_n \geq a\}$ is strictly optimal for every $0 < \alpha < 1$ – it minimizes the average detection delay $\mathbb{E}^\pi(T - \nu | T > \nu)$ if $a = a_\alpha$ is selected so that $\text{PFA}(T_a) = \alpha$ and the prior distribution is geometric. As in Tartakovsky et al. (2014); Tartakovsky and Veeravalli (2005), we refer to this procedure as the Shiryaev detection procedure in the general non-iid case too. We now show that $\Lambda_n = g_n/(1 - g_n)$, so that the Shiryaev procedure can be written as

$$T_A = \inf \{n \geq 1 : \Lambda_n \geq A\}, \quad A > 0. \quad (2.8)$$

Indeed, $g_n = \sum_{k=-\infty}^{n-1} \mathbb{P}(\nu = k | \mathbf{X}^n)$, where

$$\begin{aligned} \mathbb{P}(\nu = k | \mathbf{X}^n) &= \frac{\pi_k \prod_{j=1}^k p_\infty(X_j | \mathbf{X}^{j-1}) \prod_{i=k+1}^n p_0(X_i | \mathbf{X}^{i-1})}{\sum_{k=-\infty}^{\infty} \pi_k \prod_{j=1}^k p_\infty(X_j | \mathbf{X}^{j-1}) \prod_{i=k+1}^n p_0(X_i | \mathbf{X}^{i-1})} \\ &= \frac{\pi_k \prod_{i=k+1}^n \mathcal{L}_i}{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i + \mathbb{P}(\nu \geq n)}, \end{aligned}$$

and we obtain

$$g_n = \frac{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i}{q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=0}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i + \mathbb{P}(\nu \geq n)}.$$

Therefore,

$$\frac{g_n}{1 - g_n} = \frac{1}{\mathbb{P}(\nu \geq n)} \left(q \prod_{i=1}^n \mathcal{L}_i + \sum_{k=1}^{n-1} \pi_k \prod_{i=k+1}^n \mathcal{L}_i \right) = \Lambda_n.$$

In particular, in the popular case of zero modified geometric prior (2.2), the statistic Λ_n is

$$\Lambda_n = \frac{q}{1 - q} \prod_{i=1}^n \left(\frac{\mathcal{L}_i}{1 - \rho} \right) + \rho \sum_{k=1}^n \prod_{i=k}^n \left(\frac{\mathcal{L}_i}{1 - \rho} \right). \quad (2.9)$$

In the following, to avoid triviality, we assume that $A > q/(1 - q)$, since otherwise $T_A = 0$ with probability 1.

By Lemma 7.2.1 in Tartakovsky et al. (2014),

$$\text{PFA}(T_A) \leq 1/(1 + A) \quad \text{for every } A > q/(1 - q), \quad (2.10)$$

and therefore setting $A = A_\alpha = (1 - \alpha)/\alpha$ guarantees that $T_A \in \mathbb{C}_\alpha$.

Another popular change detection procedure is the Shiryaev–Roberts (SR) procedure (due to Shiryaev (1963) and Roberts (1966)) given by the stopping time

$$\tilde{T}_B = \inf \{n \geq 1 : R_n \geq B\}, \quad B > 0, \quad (2.11)$$

where the statistic R_n , the SR statistic, is given by

$$R_n = \sum_{k=1}^n \prod_{i=k}^n \mathcal{L}_i, \quad n \geq 0 \quad (R_0 = 0). \quad (2.12)$$

The statistic R_n can be viewed as a limit of the statistic Λ_n/ρ as $\rho \rightarrow 0$ when the prior distribution of the change point is geometric (2.2) with $q = 0$. Indeed, see (2.9).

2.4. *r*-Quick convergence versus *r*-complete convergence

Introduce the LLRs

$$Z_i = \log \frac{p_0(X_i | \mathbf{X}^{i-1})}{p_\infty(X_i | \mathbf{X}^{i-1})}, \quad \lambda_{k+n}^k = \log \frac{d\mathbf{P}_k^{k+n}}{d\mathbf{P}_\infty^{k+n}} = \sum_{i=k+1}^{k+n} Z_i, \quad n \geq 1.$$

We need the following two definitions.

Definition 2.1 (Lai (1976); Tartakovsky et al. (2014)). *Let $r > 0$. For $k = 0, 1, 2, \dots$, we say that the normalized LLR $n^{-1}\lambda_{k+n}^k$ converges r -quickly to a constant I as $n \rightarrow \infty$ under probability \mathbf{P}_k if $\mathbb{E}_k[L_k(\varepsilon)]^r < \infty$ for all $\varepsilon > 0$, where $L_k(\varepsilon) = \sup \{n \geq 1 : |n^{-1}\lambda_{k+n}^k - I| > \varepsilon\}$ ($\sup\{\emptyset\} = 0$) is the last time when $n^{-1}\lambda_{k+n}^k$ leaves the interval $[I - \varepsilon, I + \varepsilon]$.*

Definition 2.2 (Tartakovsky et al. (2014)). *Let $r > 0$. For $k = 0, 1, 2, \dots$, we say that the normalized LLR $n^{-1}\lambda_{k+n}^k$ converges r -completely to a constant I as $n \rightarrow \infty$ under probability \mathbf{P}_k if for all $\varepsilon > 0$,*

$$\sum_{n=1}^{\infty} n^{r-1} \mathbf{P}_k \left\{ \left| n^{-1}\lambda_{k+n}^k - I \right| > \varepsilon \right\} < \infty. \quad (2.13)$$

(For $r = 1$ this mode of convergence was introduced by Hsu and Robbins (1947).)

Note first that in general r -quick convergence is a stronger property than r -complete convergence. See Lemma 2.4.1 in Tartakovsky et al. (2014). More importantly, checking r -quick convergence in applications is often much more difficult than checking r -complete convergence.

In the discrete time case, Tartakovsky and Veeravalli (2005) developed a general asymptotic Bayesian theory of changepoint detection assuming that the LLR obeys the strong law of large numbers (SLLN) with some positive and finite constant I , i.e.,

$$\frac{1}{n} \lambda_{k+n}^k \xrightarrow[n \rightarrow \infty]{P_k - \text{a.s.}} I \quad \text{for all } k \in \mathbb{Z}_+, \quad (2.14)$$

with a certain rate of convergence expressed via the r -quick convergence, specifically assuming in addition that for some $r \geq 1$

$$\sum_{k=0}^{\infty} \pi_k \mathbb{E}_k [L_k(\varepsilon)]^r < \infty. \quad (2.15)$$

A similar development was performed by Baron and Tartakovsky (2006) in continuous time, assuming that

$$\int_0^{\infty} \mathbb{E}_u [L_u(\varepsilon)]^r d\Pi_u < \infty.$$

However, as we already mentioned, verification of the latter r -quick convergence condition in particular examples is not an easy task.

In Baron and Tartakovsky (2006); Tartakovsky and Veeravalli (2005), it was conjectured that all asymptotic results, including near optimality of the Shiryaev procedure (in the sense defined in (2.7)), hold if the r -quick convergence condition (2.15) is weakened into the r -complete convergence

$$\sum_{k=0}^{\infty} \pi_k \left[\sum_{n=1}^{\infty} n^{r-1} P_k \left\{ \left| n^{-1} \lambda_{k+n}^k - I \right| > \varepsilon \right\} \right] < \infty$$

(with an obvious modification in continuous time). In the following sections, we justify this conjecture.

3. Asymptotic operating characteristics and optimality of the Shiryaev procedure

In this section, we present the main results related to asymptotic optimality of the Shiryaev detection procedure in the general non-iid case as well as in the case of independent observations.

3.1. The non-iid case

The following lemma, that establishes the asymptotic lower bounds for moments of the detection delay, will be used throughout the paper. While its proof may be found in Tartakovsky and Veeravalli (2005), parts of the proof are scattered in Tartakovsky and Veeravalli (2005), and for the sake of convenience we provide a sketch of the improved version of the proof in the Appendix.

Lemma 3.1. *Let T_A be the Shiryaev changepoint detection procedure defined in (2.8). Let, for some $\mu \geq 0$, the prior distribution of the change point satisfy condition (2.3). Assume that for some positive and finite I*

$$\lim_{M \rightarrow \infty} P_k \left(\frac{1}{M} \max_{1 \leq n \leq M} \lambda_{k+n}^k \geq (1 + \varepsilon)I \right) = 0 \quad \text{for all } \varepsilon > 0 \text{ and all } k \in \mathbb{Z}_+. \quad (3.1)$$

Then, for all $m > 0$,

$$\liminf_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathcal{C}_\alpha} \mathbb{E}^\pi [(T - \nu)^m | T > \nu]}{|\log \alpha|^m} \geq \frac{1}{(I + \mu)^m}. \quad (3.2)$$

and

$$\liminf_{A \rightarrow \infty} \frac{\mathbb{E}^\pi [(T_A - \nu)^m | T_A > \nu]}{(\log A)^m} \geq \frac{1}{(I + \mu)^m}. \quad (3.3)$$

Define

$$\Upsilon_{k,r}(\varepsilon) = \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right). \quad (3.4)$$

Recall that by (2.10), $\text{PFA}(T_A) \leq (1+A)^{-1}$ for any $0 < A < q/(1-q)$, which implies that $\text{PFA}(T_{A_\alpha}) \leq \alpha$ (i.e., $T_{A_\alpha} \in \mathbb{C}_\alpha$) for any $0 < \alpha < 1-q$ if $A = A_\alpha = (1-\alpha)/\alpha$.

The following theorem is the main result in the general non-iid case, which shows that the Shiryaev detection procedure is asymptotically optimal under mild conditions for the observations and prior distributions.

Theorem 3.1. *Let T_A be the Shiryaev changepoint detection procedure defined in (2.8). Let $r \geq 1$ and let the prior distribution of the change point satisfy condition (C). Assume that for some number $0 < I < \infty$ condition (3.1) is satisfied and that the following condition holds as well*

$$\sum_{k=0}^{\infty} \pi_k \Upsilon_{k,r}(\varepsilon) < \infty \quad \text{for all } \varepsilon > 0. \quad (3.5)$$

(i) Then for all $0 < m \leq r$

$$\lim_{A \rightarrow \infty} \frac{\mathbb{E}^\pi [(T_A - \nu)^m | T_A > \nu]}{(\log A)^m} = \frac{1}{(I + \mu)^m}. \quad (3.6)$$

(ii) If $A = A_\alpha = (1-\alpha)/\alpha$, where $0 < \alpha < 1-q$, then $T_{A_\alpha} \in \mathbb{C}_\alpha$ and it is asymptotically optimal as $\alpha \rightarrow 0$ in class \mathbb{C}_α , minimizing moments of the detection delay up to order r , i.e., for all $0 < m \leq r$,

$$\lim_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi [(T - \nu)^m | T > \nu]}{\mathbb{E}^\pi [(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu]} = 1. \quad (3.7)$$

Also, the following first-order asymptotic approximations hold:

$$\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^\pi [(T - \nu)^m | T > \nu] \sim \mathbb{E}^\pi [(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu] \sim \left(\frac{|\log \alpha|}{I + \mu} \right)^m \quad \text{as } \alpha \rightarrow 0. \quad (3.8)$$

This assertion also holds if $A = A_\alpha$ is selected so that $\text{PFA}(T_{A_\alpha}) \leq \alpha$ and $\log A_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$.

In order to prove this theorem we need the following lemma, the proof of which is given in the Appendix.

Lemma 3.2. *Let $r \geq 1$ and let the prior distribution of the change point satisfy condition (C). Then for a sufficiently large A , any $0 < \varepsilon < I + \mu$ and all $k \in \mathbb{Z}_+$,*

$$\mathbb{E}_k [(T_A - k)^+]^r \leq \left(1 + \frac{\log(A/\pi_k)}{I + \mu - \varepsilon} \right)^r + r 2^{r-1} \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon - \delta \right), \quad (3.9)$$

where $\delta \rightarrow 0$ as $A \rightarrow \infty$.

If the prior distribution is geometric (2.2) with $q = 0$, i.e., $\pi_k = \rho(1-\rho)^k$, $k \in \mathbb{Z}_+$, then for any $A > 0$, any $0 < \varepsilon < I + \mu$ and all $k \in \mathbb{Z}_+$

$$\mathbb{E}_k [(T_A - k)^+]^r \leq \left(1 + \frac{\log(A/\rho)}{I + \mu - \varepsilon} \right)^r + r 2^{r-1} \sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right), \quad (3.10)$$

where $\mu = -\log(1-\rho)$.

Proof of Theorem 3.1. (i) By Lemma 3.1, under the right-tail condition (3.1) the asymptotic lower bound (3.3) holds for all $m > 0$. Thus, to establish (3.6) it suffices to show that, under the left-tail condition (3.5),

$$\limsup_{A \rightarrow \infty} \frac{\mathbb{E}^\pi[(T_A - \nu)^r | T_A > \nu]}{(\log A)^r} \leq \frac{1}{(I + \mu)^r}. \quad (3.11)$$

Let $\varepsilon_1 = \varepsilon + \delta$. By Lemma 3.2, for any $0 < \varepsilon < I + \mu$,

$$\begin{aligned} \mathbb{E}^\pi[(T_A - \nu)^r | T_A > \nu] &= \frac{\sum_{k=0}^{\infty} \pi_k \mathbb{E}_k[(T_A - k)^+]^r}{1 - \text{PFA}(T_A)} \\ &\leq \frac{\sum_{k=0}^{\infty} \pi_k \left(1 + \frac{\log(A/\pi_k)}{I + \mu - \varepsilon}\right)^r + r 2^{r-1} \sum_{k=0}^{\infty} \pi_k \Upsilon_{k,r}(\varepsilon_1)}{A/(1 + A)}, \end{aligned} \quad (3.12)$$

where we used the inequality $1 - \text{PFA}(T_A) \geq A/(1 + A)$. By conditions (3.5) and (C),

$$\sum_{k=0}^{\infty} \pi_k \Upsilon_{k,r}(\varepsilon_1) < \infty \quad \text{for any } \varepsilon_1 > 0 \quad \text{and} \quad \sum_{k=0}^{\infty} \pi_k |\log \pi_k|^r < \infty,$$

which implies that

$$\mathbb{E}^\pi[(T_A - \nu)^r | T_A > \nu] \leq \left(\frac{\log A}{I + \mu - \varepsilon}\right)^r (1 + o(1)) \quad \text{as } A \rightarrow \infty.$$

Since ε can be arbitrarily small, the upper bound (3.11) follows and the proof of (i) is complete.

(ii) Setting $A = A_\alpha = (1 - \alpha)/\alpha$ in (3.6) yields

$$\lim_{\alpha \rightarrow 0} \frac{\mathbb{E}^\pi[(T_{A_\alpha} - \nu)^r | T_{A_\alpha} > \nu]}{|\log \alpha|^r} = \frac{1}{(I + \mu)^r}, \quad (3.13)$$

which along with the lower bound (3.2) in Lemma 3.1 completes the proof of (3.7). Finally, asymptotic approximations (3.8) follow from (3.13) and (3.7). Evidently, (3.13) and (3.7), and therefore, approximations (3.8) also hold if threshold A_α is chosen so that $T_{A_\alpha} \in \mathbb{C}_\alpha$ and $\log A_\alpha \sim |\log \alpha|$ as $\alpha \rightarrow 0$. The proof is complete. \square

Theorem 3.1 implies that the Shiryaev procedure T_A is asymptotically optimal whenever the LLR converges to a constant I r -completely. Indeed, we have the following corollary.

Corollary 3.1. *Let $r \geq 1$. Let the prior distribution of the change point satisfy condition (C). Assume that for some $0 < I < \infty$*

$$\sum_{k=0}^{\infty} \pi_k \left[\sum_{n=1}^{\infty} n^{r-1} \mathbb{P}_k \left(\left| \frac{1}{n} \lambda_{k+n}^k - I \right| > \varepsilon \right) \right] < \infty \quad \text{for all } \varepsilon > 0. \quad (3.14)$$

Then (3.6), (3.7) and (3.8) hold true.

Proof. Obviously, the r -complete convergence condition (3.14) implies both conditions (3.1) and (3.5), which immediately proves the assertion of the corollary. \square

Theorem 3.1 is very general and covers, perhaps, almost all possible non-iid models as well as a large class of prior distributions. However, note that condition (C) does not include the case where μ is strictly positive, $\mu > 0$, but may go to zero, $\mu \rightarrow 0$. Indeed, in this case, the sum in (2.4) approaches infinity, and

the results of the theorem are not applicable in general. To see this, it suffices to consider the geometric prior (2.2) with $q = 0$. Then $\mu = |\log(1 - \rho)|$ and

$$\varkappa(\rho) := \sum_{k=0}^{\infty} \pi_k |\log \pi_k| = |\log \rho| + (1 - \rho) |\log(1 - \rho)| / \rho \sim |\log \rho| \quad \text{as } \rho \rightarrow 0.$$

For $r = 1$, inequality (3.12) has the form

$$\mathbb{E}^{\pi}(T_A - \nu | T_A > \nu) \leq \frac{1 + \frac{\log A + \varkappa(\rho)}{I + \mu - \varepsilon} + \sum_{k=0}^{\infty} \rho(1 - \rho)^k \Upsilon_{k,1}(\varepsilon_1)}{A/(1 + A)}.$$

Clearly, the upper bound (3.11) holds if, and only if, $\rho = \rho_A$ decays in such a way that $|\log \rho_A| = o(\log A)$. Otherwise the argument breaks down and the results are not correct.

In the next lemma, which is analogous to Lemma 3.1, we provide an asymptotic lower bound for moments of the detection delay in class \mathbb{C}_{α} when the prior distribution $\pi^{\alpha} = \{\pi_k^{\alpha}\}$ of the change point may depend on the PFA constraint α and becomes “flat” when α vanishes. Its proof is given in the Appendix.

Lemma 3.3. *Let T_A be the Shiryaev changepoint detection procedure defined in (2.8). Let the prior distribution $\pi^{\alpha} = \{\pi_k^{\alpha}\}$ of the change point satisfy condition (2.3) with $\mu > 0$ such that $\mu = \mu_{\alpha} \rightarrow 0$ as $\alpha \rightarrow 0$. Assume that for some $0 < I < \infty$ condition (3.1) holds. Then, for all $m > 0$,*

$$\liminf_{\alpha \rightarrow 0} \frac{\inf_{T \in \mathbb{C}_{\alpha}} \mathbb{E}^{\pi^{\alpha}}[(T - \nu)^m | T > \nu]}{|\log \alpha|^m} \geq \frac{1}{I^m} \quad (3.15)$$

and

$$\liminf_{A \rightarrow \infty} \frac{\mathbb{E}^{\pi^{\alpha}}[(T_A - \nu)^m | T_A > \nu]}{(\log A)^m} \geq \frac{1}{I^m}. \quad (3.16)$$

Using this lemma, we now establish asymptotic optimality of the Shiryaev procedure in the case where $\mu = \mu_{\alpha}$ approaches zero as $\alpha \rightarrow 0$.

Theorem 3.2. *Let $r \geq 1$. Suppose that the prior distribution $\pi^{\alpha} = \{\pi_k^{\alpha}\}$ of the change point ν satisfies condition (2.3) with $\mu = \mu_{\alpha} \rightarrow 0$ as $\alpha \rightarrow 0$ and that μ_{α} approaches zero at such rate that*

$$\lim_{\alpha \rightarrow 0} \frac{\sum_{k=0}^{\infty} |\log \pi_k^{\alpha}|^r \pi_k^{\alpha}}{|\log \alpha|^r} = 0. \quad (3.17)$$

Assume that for some $0 < I < \infty$ condition (3.1) and the following uniform r -complete convergence

$$\sup_{0 \leq k < \infty} \Upsilon_{k,r}(\varepsilon) < \infty \quad (3.18)$$

are satisfied. If $A = A_{\alpha}$ is so selected that $\text{PFA}(T_{A_{\alpha}}) \leq \alpha$ and $\log A_{\alpha} \sim |\log \alpha|$ as $\alpha \rightarrow 0$, in particular $A_{\alpha} = (1 - \alpha)/\alpha$, then the Shiryaev procedure $T_{A_{\alpha}}$, given by (2.8), is asymptotically optimal as $\alpha \rightarrow 0$ in class \mathbb{C}_{α} , minimizing moments of the detection delay up to order r : for all $0 < m \leq r$

$$\inf_{T \in \mathbb{C}_{\alpha}} \mathbb{E}^{\pi^{\alpha}}[(T - \nu)^m | T > \nu] \sim \mathbb{E}^{\pi^{\alpha}}[(T_{A_{\alpha}} - \nu)^m | T_{A_{\alpha}} > \nu] \sim \left(\frac{|\log \alpha|}{I} \right)^m \quad \text{as } \alpha \rightarrow 0. \quad (3.19)$$

Proof. Substituting $A = (1 - \alpha)/\alpha$ (or $\log A_{\alpha} \sim |\log \alpha|$) in inequality (3.12), we obtain

$$\mathbb{E}^{\pi^{\alpha}}[(T_A - \nu)^r | T_A > \nu] \leq \frac{\sum_{k=0}^{\infty} \pi_k^{\alpha} \left(1 + \frac{\log((1 - \alpha)/\alpha \pi_k^{\alpha})}{I + \mu_{\alpha} - \varepsilon} \right)^r + r 2^{r-1} \sup_{k \geq 0} \Upsilon_{k,r}(\varepsilon_1)}{1 - \alpha}.$$

Using conditions (3.17) and (3.18) and taking into account that $\mu_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ yields

$$\mathbb{E}^{\pi^\alpha}[(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu] \leq \left(\frac{|\log \alpha|}{I - \varepsilon} \right)^m (1 + o(1)) \quad \text{as } \alpha \rightarrow 0.$$

Since ε is an arbitrary number in $(0, I)$, we obtain the asymptotic upper bound

$$\mathbb{E}^{\pi^\alpha}[(T_{A_\alpha} - \nu)^m | T_{A_\alpha} > \nu] \leq \left(\frac{|\log \alpha|}{I} \right)^m (1 + o(1)) \quad \text{as } \alpha \rightarrow 0,$$

which along with the lower bound (3.15) in Lemma 3.3 proves (3.19). \square

Remark 3.1. *If the prior distribution is geometric (2.2), then Theorem 3.2 holds whenever the parameter $\rho = \rho_\alpha \rightarrow 0$ at a rate $|\log \rho_\alpha| = o(|\log \alpha|)$. Indeed, see the upper bound (3.10) in Lemma 3.2.*

3.2. The case of independent observations

The results of the previous subsection show that the lower bounds (3.2) and (3.3) for moments of the detection delay hold whenever the LLR process λ_{k+n}^k obeys the SLLN (2.14), since in this case condition (3.1) is satisfied. However, in general, an almost sure convergence (2.14) is not sufficient for obtaining the upper bounds, and therefore, for asymptotic optimality of the Shiryaev procedure. In fact, this condition does not even guarantee finiteness of the average delay to detection $\mathbb{E}^\pi(T_A - \nu | T_A > \nu)$, and to obtain meaningful results we need to strengthen the SLLN into the r -complete version. On the other hand, in the iid case, where conditioned on $\nu = k$ the observations X_1, \dots, X_k are iid with pre-change density $f_\infty(x)$ and X_{k+1}, X_{k+2}, \dots are iid with post-change density $f_0(x)$, the situation is dramatically different. By Theorem 4 of Tartakovsky and Veeravalli (2005), the Shiryaev procedure asymptotically (as $\alpha \rightarrow 0$) minimizes all positive moments of the detection delay in class \mathbb{C}_α if the prior distribution is geometric and the Kullback–Leibler information number

$$\mathcal{K} = \mathbb{E}_0 \lambda_1^0 = \int \log \left(\frac{f_0(x)}{f_\infty(x)} \right) d\mu(x) \quad (3.20)$$

is positive and finite.

We now extend this result to the case where observations are independent, but not necessarily identically distributed, i.e., $p_\infty(X_i | \mathbf{X}^{i-1}) = f_{\infty,i}(X_i)$ and $p_0(X_i | \mathbf{X}^{i-1}) = f_{0,i}(X_i)$ in (2.1). More generally, we may assume that the increments Z_i of the LLR $\lambda_n^k = \sum_{i=k+1}^n Z_i$ are independent, which is always the case if the observations are independent. This slight generalization is important for certain examples with dependent observations that lead to the LLR with independent increments. See, e.g., Example 5.1 in Section 5.

Theorem 3.3. *Let T_A be the Shiryaev changepoint detection procedure defined in (2.8). Let $r \geq 1$. Assume that the LLR process $\{\lambda_{k+n}^k\}_{n \geq 1}$ has independent, not necessarily identically distributed increments under \mathbb{P}_k , $k \in \mathbb{Z}_+$. Suppose that condition (3.1) holds and the following condition*

$$\lim_{n \rightarrow \infty} \mathbb{P}_k \left(\frac{1}{n} \lambda_{\ell+n}^\ell < I - \varepsilon \right) = 0 \quad \text{for all } \varepsilon > 0, \text{ all } \ell \geq k \text{ and all } k \in \mathbb{Z}_+ \quad (3.21)$$

is satisfied.

(i) *Let the prior distribution of the change point be geometric (2.2) with $q = 0$. Then relations (3.6), (3.7) and (3.8) hold true for all $m > 0$ with $\mu = |\log(1 - \rho)|$. Therefore, the Shiryaev procedure T_{A_α} minimizes asymptotically as $\alpha \rightarrow 0$ all positive moments of the detection delay in class \mathbb{C}_α .*

(ii) Let the prior distribution be geometric with the parameter $\rho = \rho_\alpha$ that depends on α and goes to zero as $\alpha \rightarrow 0$ at such rate that

$$\lim_{\alpha \rightarrow 0} \frac{\log \rho_\alpha}{\log \alpha} = 0. \quad (3.22)$$

Then relations (3.19) hold for all $m > 0$, i.e., the Shiryaev procedure is asymptotically optimal with respect to all positive moments of the detection delay.

The idea of relaxing the r -complete convergence condition by condition (3.21) is based on splitting integration, when obtaining the upper bound for the expectation $E_k[(T_A - k)^+]^r$, into a sequence of intervals (cycles) of the size $N_A \approx \log A/(I + \mu)$ and then showing that $P_k(T_A - k > \ell N_A) \leq \delta^\ell$, $\ell = 1, 2, \dots$ for some small δ under condition (3.21), using independence of the LLR increments. The details are given below.

Proof. (i) Hereafter $\lfloor x \rfloor$ denotes the largest integer less or equal to x . Let $N_A = 1 + \lfloor \log(A/\rho)/(I + \mu - \varepsilon) \rfloor$, where $\mu = |\log(1 - \rho)|$. We need only to prove that the upper bound (3.11) holds under condition (3.21). To this end, note that we have the following chain of equalities and inequalities:

$$\begin{aligned} E_k [(T_A - k)^+]^r &= \sum_{\ell=0}^{\infty} \int_{\ell N_A}^{(\ell+1)N_A} r t^{r-1} P_k(T_A - k > t) dt \\ &\leq N_A^r + \sum_{\ell=1}^{\infty} \int_{\ell N_A}^{(\ell+1)N_A} r t^{r-1} P_k(T_A - k > t) dt \\ &\leq N_A^r + \sum_{\ell=1}^{\infty} \int_{\ell N_A}^{(\ell+1)N_A} r t^{r-1} P_k(T_A - k > \ell N_A) dt \\ &= N_A^r \left(1 + \sum_{\ell=1}^{\infty} [(\ell+1)^r - \ell^r] P_k(T_A - k > \ell N_A) \right) \\ &\leq N_A^r \left(1 + \sum_{\ell=1}^{\infty} r(\ell+1)^{r-1} P_k(T_A - k > \ell N_A) \right) \\ &\leq N_A^r \left(1 + r 2^{r-1} \sum_{\ell=1}^{\infty} \ell^{r-1} P_k(T_A - k > \ell N_A) \right). \end{aligned} \quad (3.23)$$

Introduce the following notation: $A_\rho = A/\rho$, $a_\rho = \log A_\rho - N_A |\log(1 - \rho)|$,

$$R_{n,\rho} = \sum_{m=0}^{n-1} (1 - \rho)^{m-n} \exp \{ \lambda_n^m \}, \quad R_{n,\rho}^j = \sum_{m=j}^{n-1} (1 - \rho)^{m-n} \exp \{ \lambda_n^m \}, \quad n > j, \quad j = 0, 1, \dots$$

Note that $R_{n,\rho} = \Lambda_n/\rho$ (see (2.9)). Since $R_{n,\rho} \geq R_{n,\rho}^j \geq (1 - \rho)^{j-n} \exp \{ \lambda_n^j \}$ (for any $n > j$) and the

increments of λ_n^j are independent, we obtain

$$\begin{aligned}
P_k(T_A - k > \ell N_A) &= P_k(R_{n,\rho} < A_\rho \text{ for } n = 1, \dots, k + \ell N_A) \\
&\leq P_k(R_{k+nN_A,\rho} < A_\rho \text{ for } n = 1, \dots, \ell) \\
&\leq P_k\left(R_{k+nN_A,\rho}^{k+(n-1)N_A+1} < A_\rho \text{ for } n = 1, \dots, \ell\right) \\
&\leq P_k\left(\exp\left\{\lambda_{k+nN_A}^{k+(n-1)N_A+1}\right\} < A_\rho(1-\rho)^{N_A} \text{ for } n = 1, \dots, \ell\right) \\
&= P_k\left(\lambda_{k+N_A}^k < a_\rho, \lambda_{k+2N_A}^{k+N_A+1} < a_\rho, \dots, \lambda_{k+\ell N_A}^{k+(\ell-1)N_A+1} < a_\rho\right) \\
&= \prod_{n=1}^{\ell} P_k\left(\lambda_{k+nN_A}^{k+(n-1)N_A+1} < a_\rho\right) \leq \prod_{n=1}^{\ell} P_k\left(\frac{1}{N_A} \lambda_{k+nN_A}^{k+(n-1)N_A+1} < I - \varepsilon\right),
\end{aligned} \tag{3.24}$$

where the last inequality follows from the inequality

$$\begin{aligned}
P_k\left(\lambda_{k+nN_A}^{k+(n-1)N_A+1} < a_\rho\right) &= P_k\left(\frac{\lambda_{k+nN_A}^{k+(n-1)N_A+1}}{N_A} < \frac{\log A_\rho}{N_A} - \mu\right) \\
&\leq P_k\left(\frac{\lambda_{k+nN_A}^{k+(n-1)N_A+1}}{N_A} < \frac{\log A_\rho}{1 + \log A_\rho}(I + \mu - \varepsilon) - \mu\right) \\
&\leq P_k\left(\frac{\lambda_{k+nN_A}^{k+(n-1)N_A+1}}{N_A} < I - \varepsilon\right),
\end{aligned}$$

which holds for all $0 < \varepsilon < I + \mu$ and $k \in \mathbb{Z}_+$. By condition (3.21), for a sufficiently large A there exists a small δ_A such that

$$P_k\left(\frac{1}{N_A} \lambda_{k+nN_A}^{k+(n-1)N_A+1} < I - \varepsilon\right) \leq \delta_A, \quad n \geq 1.$$

Therefore, for any $\ell \geq 1$,

$$P_k(T_A - k > \ell N_A) \leq \delta_A^\ell.$$

Combining this inequality with (3.23) and using the fact that $L_{r,A} = \sum_{\ell=1}^{\infty} \ell^{r-1} \delta_A^\ell \rightarrow 0$ as $A \rightarrow \infty$ for any $r > 0$, we obtain

$$\begin{aligned}
E^\pi[(T_A - \nu)^r | T_A > \nu] &= \frac{\sum_{k=0}^{\infty} \pi_k E_k[(T_A - k)^+]^r}{1 - \text{PFA}(T_A)} \\
&\leq \frac{\left(1 + \frac{\log(A/\rho)}{I + \mu - \varepsilon}\right)^r + r 2^{r-1} L_{r,A}}{A/(1+A)} \\
&= \left(\frac{\log A}{I + \mu - \varepsilon}\right)^r (1 + o(1)) \quad \text{as } A \rightarrow \infty.
\end{aligned} \tag{3.25}$$

Since $\varepsilon \in (0, I + \mu)$ is an arbitrary number this implies the upper bound (3.11).

Applying (3.11) together with the lower bound (3.3) (which holds as before due to condition (3.1)) yields (3.6).

Next, under condition (3.1), for all $m > 0$, we have the asymptotic lower bound (3.2) in class \mathbb{C}_α . Substituting $\log A_\alpha \sim |\log \alpha|$ in (3.6) (in particular, we may take $A_\alpha = (1 - \alpha)/\alpha$) we immediately obtain the asymptotic approximation (3.13) for moments of the detection delay of the Shiryaev procedure T_{A_α} . This proves (3.7). Asymptotic approximations (3.8) are obvious from (3.13) and (3.7). This completes the proof of (i).

(ii) Substituting $\rho = \rho_\alpha$ and $A = A_\alpha = (1 - \alpha)/\alpha$ (or more generally $\log A_\alpha \sim |\log \alpha|$) in inequality (3.25), we obtain

$$\mathbb{E}^{\pi^\alpha}[(T_{A_\alpha} - \nu)^r | T_{A_\alpha} > \nu] \leq \frac{\left(1 + \frac{\log((1-\alpha)/\alpha) + |\log \rho_\alpha|}{I + \mu_\alpha - \varepsilon}\right)^r + r2^{r-1} L_{r, A_\alpha}}{1 - \alpha}.$$

By condition (3.22), the right side is asymptotically as $\alpha \rightarrow 0$ equal to

$$\left(\frac{|\log \alpha|}{I - \varepsilon}\right)^r (1 + o(1)),$$

which along with the lower bound (3.2) ($\mu_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$) completes the proof. \square

Remark 3.2. The assertions of Theorem 3.3 hold whenever the normalized LLR processes $n^{-1}\lambda_{\ell+n}^\ell$, $\ell = k, k+1, \dots$ converge almost surely to a constant I under \mathbb{P}_k for all $k \in \mathbb{Z}_+$, since in this case both conditions (3.1) and (3.21) are satisfied. In the iid case, the assertions of the theorem are true with $I = \mathcal{K}$ being the Kullback–Leibler information number (3.20), assuming that $0 < \mathcal{K} < \infty$. Indeed, in this case, the conditions of Theorem 3.3 hold with $I = \mathcal{K}$ by the SLLN. This result has been previously established by Tartakovsky and Veeravalli (2005) using a completely different technique.

Remark 3.3. Theorem 3.3(i) can be generalized for the arbitrary prior distribution satisfying condition (C) and Theorem 3.3(ii) for prior distributions satisfying condition (2.3) with parameter $\mu = \mu_\alpha = o(|\log \alpha|)$ as $\alpha \rightarrow 0$. However, in this general case, the proof becomes very tedious and obscures the main ideas. For this reason, we focused on the geometric prior, which is not an overly restrictive assumption, especially in part (ii).

4. Asymptotic operating characteristics of the Shiryaev–Roberts procedure

In this section, we discuss asymptotic operating characteristics of the SR procedure \tilde{T}_B defined in (2.11) and (2.12). While the methods are similar to those used in the previous section, there are specific features and certain differences that have to be considered separately.

4.1. The non-iid case

As mentioned in Subsection 2.3, the SR statistic R_n is the limit of the statistic Λ_n/ρ as $\rho \rightarrow 0$ when the prior distribution is geometric (2.2) with $q = 0$. Therefore, it is intuitively appealing, based on the results of Theorem 3.2, that

$$\mathbb{E}^\pi(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu \sim \left(\frac{\log B}{I}\right)^r \quad \text{as } B \rightarrow \infty,$$

and therefore, if we can select $B = B_\alpha$ so that $\text{PFA}(\tilde{T}_{B_\alpha}) \leq \alpha$ and $\log B_\alpha \sim |\log \alpha|$, then the SR procedure is also asymptotically as $\alpha \rightarrow 0$ optimal whenever $\rho_\alpha \rightarrow 0$ at an appropriate rate. Below we show that this is indeed true.

The first question is how to select threshold B_α in order to embed the SR procedure into class \mathbb{C}_α . To answer this question, it suffices to note that under \mathbb{P}_∞ the SR statistic R_n is a submartingale with mean $\mathbb{E}_\infty R_n = n$, so that applying Doob's submartingale inequality, we obtain

$$\mathbb{P}_\infty(\tilde{T}_B \leq j) = \mathbb{P}_\infty\left(\max_{1 \leq i \leq j} R_i \geq B\right) \leq j/B, \quad j = 1, 2, \dots,$$

and hence,

$$\text{PFA}(\tilde{T}_B) = \sum_{j=0}^{\infty} \pi_j \mathbb{P}_{\infty}(\tilde{T}_B \leq j) \leq \bar{\nu}/B, \quad (4.1)$$

where $\bar{\nu} = \sum_{j=1}^{\infty} j\pi_j$. Therefore, assuming that $\bar{\nu} < \infty$, we obtain that setting $B = B_{\alpha} = \bar{\nu}/\alpha$ implies $\tilde{T}_{B_{\alpha}} \in \mathbb{C}_{\alpha}$. If, in a particular case, the prior distribution is geometric, then $\text{PFA}(\tilde{T}_B) \leq (1 - \rho)/(\rho B)$.

Theorem 4.1. *Let \tilde{T}_B be the SR changepoint detection procedure defined in (2.11). Let $\bar{\nu} = \sum_{j=1}^{\infty} j\pi_j < \infty$. Let $r \geq 1$. Assume that for some number $0 < I < \infty$ conditions (3.1) and (3.5) are satisfied.*

(i) *Then for all $0 < m \leq r$*

$$\lim_{B \rightarrow \infty} \frac{\mathbb{E}^{\pi}[(\tilde{T}_B - \nu)^m | \tilde{T}_B > \nu]}{(\log B)^m} = \frac{1}{I^m}. \quad (4.2)$$

(ii) *Let $B = B_{\alpha} = \bar{\nu}/\alpha$. Then $\tilde{T}_{B_{\alpha}} \in \mathbb{C}_{\alpha}$ and for all $0 < m \leq r$,*

$$\lim_{\alpha \rightarrow 0} \frac{\mathbb{E}^{\pi}[(\tilde{T}_{B_{\alpha}} - \nu)^m | \tilde{T}_{B_{\alpha}} > \nu]}{|\log \alpha|^m} = \frac{1}{I^m}. \quad (4.3)$$

This assertion also holds if $B = B_{\alpha}$ is selected so that $\text{PFA}(T_{B_{\alpha}}) \leq \alpha$ and $\log B_{\alpha} \sim |\log \alpha|$ as $\alpha \rightarrow 0$.

Proof. (i) For $\varepsilon \in (0, 1)$, let $M_{B,\varepsilon} = (1 - \varepsilon)I^{-1} \log B$. Using Chebyshev's inequality and inequality (4.1), similarly to (A.1) we obtain that

$$\mathbb{E}^{\pi}[(\tilde{T}_B - \nu)^m | \tilde{T}_B > \nu] \geq M_{B,\varepsilon}^m \left[1 - \bar{\nu}/B - \mathbb{P}^{\pi}(0 < \tilde{T}_B - \nu < M_{B,\varepsilon}) \right]. \quad (4.4)$$

Now, similarly to (A.3),

$$\mathbb{P}_k(0 < \tilde{T}_B - k < M_{B,\varepsilon}) \leq U_{B,\varepsilon}^k(\tilde{T}_B) + \beta_{B,\varepsilon}^k, \quad (4.5)$$

where

$$U_{B,\varepsilon}^k(\tilde{T}_B) = e^{(1+\varepsilon)IM_{B,\varepsilon}} \mathbb{P}_{\infty}(0 < \tilde{T}_B - k < M_{B,\varepsilon}), \quad \beta_{B,\varepsilon}^k = \mathbb{P}_k\left(\frac{1}{M_{B,\varepsilon}} \max_{1 \leq n \leq M_{B,\varepsilon}} \lambda_{k+n}^k \geq (1 + \varepsilon)I\right).$$

Since

$$\mathbb{P}_{\infty}(0 < \tilde{T}_B - k < M_{B,\varepsilon}) \leq \mathbb{P}_{\infty}(\tilde{T}_B < k + M_{B,\varepsilon}) \leq (k + M_{B,\varepsilon})/B,$$

we have

$$U_{B,\varepsilon}^k(\tilde{T}_B) \leq \frac{k + (1 - \varepsilon)I^{-1} \log B}{B\varepsilon^2}. \quad (4.6)$$

Let K_B be an integer number that approaches infinity as $B \rightarrow \infty$. Using (4.5) and (4.6), we obtain the following upper bound

$$\begin{aligned} \mathbb{P}^{\pi}(0 < \tilde{T}_B - \nu < M_{B,\varepsilon}) &= \sum_{k=0}^{\infty} \pi_k \mathbb{P}_k(0 < \tilde{T}_B - k < M_{B,\varepsilon}) \\ &\leq \mathbb{P}(\nu > K_B) + \sum_{k=0}^{\infty} \pi_k U_{B,\varepsilon}^k(\tilde{T}_B) + \sum_{k=0}^{K_B} \pi_k \beta_{B,\varepsilon}^k \\ &\leq \mathbb{P}(\nu > K_B) + \frac{\bar{\nu} + (1 - \varepsilon)I^{-1} \log B}{B\varepsilon^2} + \sum_{k=0}^{K_B} \pi_k \beta_{B,\varepsilon}^k, \end{aligned} \quad (4.7)$$

where the first two terms go to zero as $B \rightarrow \infty$ for all $\varepsilon > 0$ since $\bar{\nu}$ is finite (note that by Markov's inequality $P(\nu > K_B) \leq \bar{\nu}/K_B$) and the last term also goes to zero by condition (3.1) and Lebesgue's dominated convergence theorem. Thus, for all $0 < \varepsilon < 1$,

$$P^\pi(0 < \tilde{T}_B - \nu < M_{B,\varepsilon}) \rightarrow 0 \quad \text{as } B \rightarrow \infty$$

and applying inequality (4.4), we obtain that for any $0 < \varepsilon < 1$ as $B \rightarrow \infty$

$$E^\pi[(\tilde{T}_B - \nu)^m | \tilde{T}_B > \nu] \geq (1 - \varepsilon)^m \left(\frac{\log B}{I} \right)^m (1 + o(1)).$$

Since ε can be arbitrarily small, this inequality yields the asymptotic lower bound (for any $m > 0$)

$$E^\pi[(\tilde{T}_B - \nu)^m | \tilde{T}_B > \nu] \geq \left(\frac{\log B}{I} \right)^m (1 + o(1)) \quad \text{as } B \rightarrow \infty. \quad (4.8)$$

To complete the proof of assertion (i), we now need to show that this lower bound is also an upper bound asymptotically as $B \rightarrow \infty$. In just the same way as in the proof of Lemma 3.2 (see (A.5)), we obtain

$$\begin{aligned} E_k \left[(\tilde{T}_B - k)^+ \right]^r &= \int_0^\infty r t^{r-1} P_k(\tilde{T}_B - k > t) dt \\ &\leq N_{B,\varepsilon}^r + r 2^{r-1} \sum_{n=N_{B,\varepsilon}}^\infty n^{r-1} P_k(\tilde{T}_B - k > n), \end{aligned} \quad (4.9)$$

where $N_{B,\varepsilon} = 1 + \lfloor (\log B)/(I - \varepsilon) \rfloor$. Clearly, $R_n \geq e^{\lambda_n^k}$ (for any $n > k$), and therefore,

$$P_k(\tilde{T}_B - k > n) = P_k\left(\max_{1 \leq i \leq n+k} R_i < B\right) \leq P_k\left(\frac{1}{n} \lambda_{k+n}^k < \frac{1}{n} \log B\right).$$

But for all $k \in \mathbb{Z}_+$ and $n \geq N_{B,\varepsilon}$ the latter probability can be upper-bounded as

$$P_k\left(\frac{1}{n} \lambda_{k+n}^k < \frac{1}{n} \log B\right) \leq P_k\left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon\right),$$

so that for all $k \in \mathbb{Z}_+$ and $n \geq N_{B,\varepsilon}$

$$P_k(\tilde{T}_B - k > n) \leq P_k\left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon\right).$$

Substituting this upper bound in inequality (4.9) yields (for every $0 < \varepsilon < I$)

$$\begin{aligned} E^\pi[(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu] &= \frac{\sum_{k=0}^\infty \pi_k E_k[(\tilde{T}_B - k)^+]^r}{1 - \text{PFA}(\tilde{T}_B)} \\ &\leq \frac{\left(1 + \frac{\log B}{I - \varepsilon}\right)^r + r 2^{r-1} \sum_{k=0}^\infty \pi_k \Upsilon_{k,r}(\varepsilon)}{1 - \bar{\nu}/B}, \end{aligned}$$

where we used the inequality $1 - \text{PFA}(\tilde{T}_B) \geq 1 - \bar{\nu}/B$. By condition (3.5),

$$\sum_{k=0}^\infty \pi_k \Upsilon_{k,r}(\varepsilon) < \infty \quad \text{for any } \varepsilon > 0,$$

which implies that, for every $0 < \varepsilon < I$

$$\mathbb{E}^\pi[(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu] \leq \left(\frac{\log B}{I - \varepsilon} \right)^r (1 + o(1)) \quad \text{as } B \rightarrow \infty.$$

Since ε can be arbitrarily small, this implies the asymptotic upper bound

$$\mathbb{E}^\pi[(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu] \leq \left(\frac{\log B}{I} \right)^r (1 + o(1)) \quad \text{as } B \rightarrow \infty. \quad (4.10)$$

This completes the proof of (i).

(ii) Substitution of $B = B_\alpha = \bar{\nu}/\alpha$ (or more generally $\log B_\alpha \sim |\log \alpha|$) in (4.2) immediately yields (4.3), and the proof is complete. \square

Theorem 4.1 does not cover the case of prior distributions with exponential tails ($\mu > 0$) but with $\mu = \mu_\alpha$ that depends on the PFA α and approaches zero as $\alpha \rightarrow 0$. The next theorem, which is similar to Theorem 3.2, addresses this case.

Theorem 4.2. *Let $r \geq 1$. Assume that the prior distribution $\pi^\alpha = \{\pi_k^\alpha\}$ of the change point ν satisfies condition (2.3) with $\mu = \mu_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ and that μ_α approaches zero at such rate that $\bar{\nu}_\alpha = \sum_{k=1}^\infty k \pi_k^\alpha$ increases at a rate slower than $|\log \alpha|$ as $\alpha \rightarrow 0$, i.e.,*

$$\lim_{\alpha \rightarrow 0} \frac{\bar{\nu}_\alpha}{|\log \alpha|} = 0. \quad (4.11)$$

Assume that for some $0 < I < \infty$ conditions (3.1) and (3.18) are satisfied. If $B = B_\alpha = \bar{\nu}_\alpha/\alpha$, then $\text{PFA}(T_{B_\alpha}) \leq \alpha$ and for all $0 < m \leq r$

$$\inf_{T \in \mathbb{C}_\alpha} \mathbb{E}^{\pi^\alpha}[(T - \nu)^m | T > \nu] \sim \mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^m | \tilde{T}_{B_\alpha} > \nu] \sim \left(\frac{|\log \alpha|}{I} \right)^m \quad \text{as } \alpha \rightarrow 0. \quad (4.12)$$

Therefore, the SR procedure \tilde{T}_{B_α} is asymptotically optimal as $\alpha \rightarrow 0$ in class \mathbb{C}_α , minimizing moments of the detection delay up to order r .

Proof. Similarly to (4.4),

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^m | \tilde{T}_{B_\alpha} > \nu] \geq M_{B_\alpha, \varepsilon}^m \left[1 - \alpha - \mathbb{P}^{\pi^\alpha} \left(0 < \tilde{T}_{B_\alpha} - \nu < M_{B_\alpha, \varepsilon} \right) \right].$$

Let $K_\alpha = \lfloor C |\log \alpha| \rfloor$ with some positive constant C . Substituting $B = B_\alpha = \bar{\nu}_\alpha/\alpha$ in inequality (4.7), we obtain

$$\mathbb{P}^{\pi^\alpha} (0 < \tilde{T}_{B_\alpha} - \nu < M_{B_\alpha, \varepsilon}) \leq \mathbb{P}(\nu > K_\alpha) + \frac{\bar{\nu}_\alpha + (1 - \varepsilon)I^{-1} \log(\bar{\nu}_\alpha/\alpha)}{(\bar{\nu}_\alpha/\alpha)^{\varepsilon^2}} + \sum_{k=0}^{K_\alpha} \pi_k^\alpha \beta_{B_\alpha, \varepsilon}^k.$$

As before, the last term approaches zero as $\alpha \rightarrow 0$. It is easily verified that the middle term also approaches zero as long as condition (4.11) is satisfied. Finally, by the Markov inequality and condition (4.11),

$$\mathbb{P}(\nu > K_\alpha) \leq \bar{\nu}_\alpha/K_\alpha = o(|\log \alpha|)/C|\log \alpha| \rightarrow 0 \quad \text{as } \alpha \rightarrow 0,$$

so $\lim_{\alpha \rightarrow 0} \mathbb{P}(\nu > K_\alpha) = 0$. Therefore, we conclude that for all $0 < \varepsilon < 1$,

$$\mathbb{P}^\pi(0 < \tilde{T}_{B_\alpha} - \nu < M_{B_\alpha, \varepsilon}) \rightarrow 0 \quad \text{as } \alpha \rightarrow 0.$$

Since by (4.11), $\log B_\alpha \sim |\log \alpha|$, we obtain that, for all $m > 0$,

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^m | \tilde{T}_{B_\alpha} > \nu] \geq \left(\frac{|\log \alpha|}{I} \right)^m (1 + o(1)) \quad \text{as } \alpha \rightarrow 0. \quad (4.13)$$

The upper bound

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^m | \tilde{T}_{B_\alpha} > \nu] \leq \left(\frac{|\log \alpha|}{I} \right)^m (1 + o(1)) \quad \text{as } \alpha \rightarrow 0 \quad (4.14)$$

is obtained in the manner absolutely analogous to the proof of the upper bound (4.10) in the previous theorem. Specifically, for all $k \in \mathbb{Z}_+$ and $n \geq N_{B_\alpha, \varepsilon}$

$$\mathbb{P}_k(\tilde{T}_{B_\alpha} - k > n) \leq \mathbb{P}_k\left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon\right)$$

and

$$\mathbb{E}_k[(\tilde{T}_B - k)^+]^r \leq N_{B_\alpha, \varepsilon}^r + r2^{r-1} \sum_{n=N_{B_\alpha, \varepsilon}}^{\infty} n^{r-1} \mathbb{P}_k(\tilde{T}_{B_\alpha} - k > n),$$

so that

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^r | \tilde{T}_{B_\alpha} > \nu] \leq \frac{\left(1 + \frac{\log(\bar{\nu}_\alpha/\alpha)}{I - \varepsilon}\right)^r + r2^{r-1} \sup_{k \geq 0} \Upsilon_{k,r}(\varepsilon)}{1 - \alpha}.$$

Taking into account that, by condition (4.11), $\log(\bar{\nu}_\alpha/\alpha) \sim |\log \alpha|$ and that, by condition (3.18), $\sup_{k \geq 0} \Upsilon_{k,r}(\varepsilon) < \infty$, yields the asymptotic upper bound (4.14).

Now, applying the bounds (4.13) and (4.14) simultaneously, we obtain the asymptotic approximation

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^m | \tilde{T}_{B_\alpha} > \nu] \sim \left(\frac{|\log \alpha|}{I} \right)^m \quad \text{as } \alpha \rightarrow 0,$$

i.e., the second approximation in (4.12). The first one follows from the lower bound (3.2) (with $\mu = \mu_\alpha \rightarrow 0$). The proof is complete. \square

Remark 4.1. If the prior distribution is geometric (2.2), then $\mu = |\log(1 - \rho)|$ and $\bar{\nu} = (1 - \rho)/\rho$ and Theorem 4.2 holds whenever the parameter $\rho = \rho_\alpha \rightarrow 0$ at the rate $|\log \rho_\alpha| = o(|\log \alpha|)$.

Comparing asymptotic formula (4.3) with asymptotic lower bound (3.2) in class \mathbb{C}_α , we see that, opposed to the Shiryaev procedure, the SR procedure is not asymptotically optimal so long as $\mu > 0$, i.e., if the prior distribution has exponential tail. If the tail is heavy, i.e., $\mu = 0$, then the SR procedure is asymptotically optimal and, by Theorem 4.2, the same is true if $\mu = \mu_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ at a suitable rate. This is intuitively expected, since the SR statistic does not exploit the prior distribution, relying on the improper uniform prior. However, there still may be a problem when applying the latter asymptotic result in practice. Indeed, there is no guarantee that the bound $\bar{\nu}/B$ in inequality (4.1) is relatively tight in a sense that $|\log \text{PFA}(\tilde{T}_B)| \sim \log B$ as $B \rightarrow \infty$, i.e., that for a sufficiently large B , $\text{PFA}(\tilde{T}_B) \approx \text{const}/B$, unless μ/I is small if the prior satisfies condition (2.3) with $\mu > 0$. Even in the case of heavy-tailed priors ($\mu = 0$) this is perhaps not true. In this respect, Tartakovsky and Moustakides (2010) conjectured that asymptotically as $B \rightarrow \infty$ the accurate approximation is $\text{PFA}(\tilde{T}_B) \sim O(1)/B^{s(\mu)}$, where $s(\mu) > 1$ for all $\mu > 0$ and $s(\mu) \rightarrow 0$ as $\mu \rightarrow 0$. If this conjecture is correct, which is partially justified in Tartakovsky and Moustakides (2010) by numerical computations, then the asymptotic relative efficiency of the asymptotically optimal Shiryaev procedure compared to the SR procedure is $[Is(\mu)/(I + \mu)]^m$, but not $[I/(I + \mu)]^m$, as Theorem 4.1 suggests. Note that this is expected to be true only for the priors with the exponential tail, but not necessarily for heavy-tailed priors.

4.2. The case of independent observations

We now provide a theorem for the SR procedure similar to Theorem 3.3 in the case where the LLR has independent increments.

Theorem 4.3. *Let \tilde{T}_B be the SR changepoint detection procedure. Let $r \geq 1$. Assume that the LLR process $\{\lambda_{k+n}^k\}_{n \geq 1}$ has independent, not necessarily identically distributed increments under P_k , $k \in \mathbb{Z}_+$. Suppose that conditions (3.1) and (3.21) are satisfied.*

(i) *Let the prior distribution $\pi = \{\pi_k\}$ be geometric (2.2) with $q = 0$. Then relation (4.2) holds for all $m > 0$. If $B = B_\alpha = (1 - \rho)/\rho\alpha$, then $\tilde{T}_{B_\alpha} \in \mathbb{C}_\alpha$ and relation (4.3) holds for all $m > 0$.*

(ii) *Let the prior distribution $\pi^\alpha = \{\pi_k^\alpha\}$ be geometric with the parameter $\rho = \rho_\alpha$ that depends on α and goes to zero as $\alpha \rightarrow 0$ at such a rate that condition (3.22) is satisfied. Then relations (4.12) hold for all $m > 0$, i.e., the SR procedure is asymptotically optimal with respect to all positive moments of the detection delay.*

Proof. (i) Again let $N_{B,\varepsilon} = 1 + \lfloor \log(B)/(I - \varepsilon) \rfloor$. Similarly to (3.23), we obtain

$$\mathbb{E}_k \left[(\tilde{T}_B - k)^+ \right]^r \leq N_{B,\varepsilon}^r \left(1 + r2^{r-1} \sum_{\ell=1}^{\infty} \ell^{r-1} P_k(\tilde{T}_B - k > \ell N_{B,\varepsilon}) \right). \quad (4.15)$$

Since

$$R_n \geq R_n^j = \sum_{m=j}^{n-1} e^{\lambda_n^m} \geq e^{\lambda_n^j}, \quad n > j, \quad j = 0, 1, \dots,$$

and the increments of λ_n^j are independent, as in (3.24), we obtain that for all $0 < \varepsilon < I$ and $k \in \mathbb{Z}_+$

$$P_k(\tilde{T}_B - k > \ell N_{B,\varepsilon}) \leq \prod_{n=1}^{\ell} P_k(\lambda_{k+nN_{B,\varepsilon}}^{k+(n-1)N_{B,\varepsilon}+1} < \log B) \leq \prod_{n=1}^{\ell} P_k\left(\frac{1}{N_{B,\varepsilon}} \lambda_{k+nN_{B,\varepsilon}}^{k+(n-1)N_{B,\varepsilon}+1} < I - \varepsilon\right).$$

By condition (3.21), for a sufficiently large B there exists a small δ_B such that

$$P_k\left(\frac{1}{N_{B,\varepsilon}} \lambda_{k+nN_{B,\varepsilon}}^{k+(n-1)N_{B,\varepsilon}+1} < I - \varepsilon\right) \leq \delta_B, \quad n \geq 1,$$

so that, for any $\ell \geq 1$, $P_k(\tilde{T}_B - k > \ell N_{B,\varepsilon}) \leq \delta_B^\ell$. This along with (3.23) and the fact that $L_{r,B} = \sum_{\ell=1}^{\infty} \ell^{r-1} \delta_B^\ell \rightarrow 0$ as $B \rightarrow \infty$ for any $r > 0$ yields

$$\begin{aligned} \mathbb{E}^\pi[(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu] &\leq \frac{\left(1 + \frac{\log B}{I - \varepsilon}\right)^r + r2^{r-1} L_{r,B}}{1 - \bar{\nu}/B} \\ &= \left(\frac{\log B}{I - \varepsilon}\right)^r (1 + o(1)) \quad \text{as } B \rightarrow \infty. \end{aligned} \quad (4.16)$$

Since ε is arbitrarily small, it follows that

$$\mathbb{E}^\pi[(\tilde{T}_B - \nu)^r | \tilde{T}_B > \nu] \leq \left(\frac{\log B}{I}\right)^r (1 + o(1)) \quad \text{as } B \rightarrow \infty. \quad (4.17)$$

Since the lower bound (4.8) holds even in a more general case this proves (4.2).

Next, under condition (3.1) the asymptotic lower bound (4.13) holds (for all $m > 0$) in class \mathbb{C}_α even in a more general case (see the proof of Theorem 4.2). Substituting in (4.17) $B_\alpha = (1 - \rho)/\rho\alpha$ (or more generally $\log B_\alpha \sim |\log \alpha|$), we obtain (4.3), which completes the proof of (i).

(ii) Substituting $B = B_\alpha = (1 - \rho_\alpha)/\rho_\alpha\alpha$ (or more generally $\log B_\alpha \sim |\log \alpha|$) in inequality (4.17), we obtain

$$\mathbb{E}^{\pi^\alpha}[(\tilde{T}_{B_\alpha} - \nu)^r | \tilde{T}_{B_\alpha} > \nu] \leq \left(\frac{|\log \alpha|}{I} \right)^r (1 + o(1)).$$

This upper bound along with the lower bound (3.2) with $\mu_\alpha \rightarrow 0$ as $\alpha \rightarrow 0$ proves asymptotic relations (4.12). \square

Remark 4.2. Both conditions (3.1) and (3.21) are satisfied when the normalized LLRs $n^{-1}\lambda_{\ell+n}^\ell$, $\ell = k, k+1, \dots$ converge a.s. to I under \mathbb{P}_k . Therefore, the SLLN is sufficient for Theorem 4.3, i.e., for asymptotic optimality of the SR procedure with respect to all positive moments of the detection delay.

5. Examples

We now consider three examples that illustrate the general asymptotic theory developed in previous sections.

Example 5.1 (Detection of a deterministic signal in AR noise). Let S_n be a deterministic function (signal) that appears at an unknown time ν in additive noise ξ_n , so the observations have the standard “signal-plus-noise/clutter” form

$$X_n = \mathbb{1}_{\{n > \nu\}} S_n + \xi_n, \quad n \geq 1,$$

where $\{\xi_n\}_{n \in \mathbb{Z}_+}$ is a p -th order autoregression (AR(p) process) driven by the normal $\mathcal{N}(0, \sigma^2)$ iid sequence $\{w_n\}_{n \geq 1}$, i.e., the sequence $\{\xi_n\}_{n \geq 1}$ obeys the recursion

$$\xi_n = \sum_{i=1}^p \beta_i \xi_{n-i} + w_n, \quad n \geq 1, \quad \xi_{1-p} = \xi_{2-p} = \dots = \xi_0 = 0. \quad (5.1)$$

The coefficients β_1, \dots, β_p and variance σ^2 are known, and we suppose that $\beta_1 + \dots + \beta_p \neq 1$. Let $\varphi(x) = (2\pi)^{-1/2} e^{-x^2/2}$ denote density of the standard normal distribution. Define the p -th order residual

$$\tilde{X}_n = \begin{cases} X_n - \sum_{i=1}^p \beta_i X_{n-i} & \text{for } n > p \\ X_n - \sum_{i=1}^{j-1} \beta_i X_{j-i} & \text{for } 1 \leq n = j \leq p \end{cases}.$$

It is easy to see that pre- and post-change conditional densities $p_\infty(X_n | \mathbf{X}^{n-1})$ and $p_0(X_n | \mathbf{X}^{n-1})$ are

$$p_\infty(X_n | \mathbf{X}^{n-1}) = \frac{1}{\sigma} \varphi\left(\frac{\tilde{X}_n}{\sigma}\right), \quad p_0(X_n | \mathbf{X}^{n-1}) = \frac{1}{\sigma} \varphi\left(\frac{\tilde{X}_n - \tilde{S}_n}{\sigma}\right), \quad (5.2)$$

where

$$\tilde{S}_n = S_n - \sum_{i=1}^p \beta_i S_{n-i} \quad \text{for } n > p \quad \text{and} \quad \tilde{S}_n = S_n - \sum_{i=1}^{j-1} \beta_i S_{j-i} \quad \text{for } 1 \leq n = j \leq p.$$

Using (5.2), we obtain that for all $k \in \mathbb{Z}_+$ and $n \geq 1$ the LLR has the form

$$\lambda_{k+n}^k = \frac{1}{\sigma^2} \sum_{j=k+1}^{k+n} \tilde{S}_j \tilde{X}_j - \frac{1}{2\sigma^2} \sum_{j=k+1}^{k+n} \tilde{S}_j^2.$$

Thus, the initial problem of detection of the signal S_n that appears at unknown time ν in the correlated AR noise reduces to detection of the transformed signal \tilde{S}_n in white Gaussian noise. As a result, the LLR has independent (but not identically distributed) increments. Since under measure \mathbb{P}_k the random variables

$\{\tilde{X}_n\}_{n \geq k+1}$ are independent Gaussian random variables with mean $E_k \tilde{X}_n = \tilde{S}_n$ and variance σ^2 , under P_k the LLR can be represented as

$$\lambda_{k+n}^k = \frac{1}{2\sigma^2} \sum_{j=k+1}^{k+n} \tilde{S}_j^2 + \frac{1}{\sigma} \sum_{j=k+1}^{k+n} \tilde{S}_j \eta_j,$$

where $\eta_j, j \geq k+1$ are iid standard normal random variables, $\eta_j \sim \mathcal{N}(0, 1)$.

Assume that the energy of the transformed signal is an asymptotically linear function, i.e.,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=k+1}^{k+n} \tilde{S}_j^2 = \tilde{S}^2, \quad 0 < \tilde{S}^2 < \infty.$$

Then for all $k \in \mathbb{Z}_+$

$$\frac{1}{n} \lambda_{k+n}^k \xrightarrow[n \rightarrow \infty]{P_k\text{-a.s.}} \frac{\tilde{S}^2}{2\sigma^2} = I$$

and, by Theorem 3.3, the Shiryaev procedure minimizes as $\alpha \rightarrow 0$ all positive moments of the detection delay for any value of the parameter $0 < \rho < 1$ of the geometric prior. By Theorem 4.3, the SR procedure also minimizes all moments of the detection delay if $\rho = \rho_\alpha \rightarrow 0$ and $|\log \rho_\alpha| = o(|\log \alpha|)$ as $\alpha \rightarrow 0$. If $S_n = S$ does not depend on n , then

$$I = \frac{S^2}{2\sigma^2} \left(1 - \sum_{i=1}^p \beta_i \right)^2.$$

Example 5.2 (Detection of a change of variance in normal population with unknown mean). Let observations $X_n \sim \mathcal{N}(\theta, \sigma_\infty^2)$ be iid normal with variance σ_∞^2 before change and iid normal $\mathcal{N}(\theta, \sigma_0^2)$ with variance σ_0^2 after change with the same unknown mean θ . Formally, this problem is not in the class of problems considered in this paper since pre- and post-change densities depend on an unknown nuisance parameter θ , and hence, the hypotheses are not simple, but composite. However, this problem can be reduced to the problem of testing simple hypotheses using the principle of invariance, since it is invariant under the group of shifts $\{G_b(x) = x + b\}_{-\infty < b < \infty}$. The maximal invariant is $\mathbf{Y}^n = (Y_1, \dots, Y_n)$, $n \geq 2$, where $Y_k = X_k - X_1$, $Y_1 = 0$, and we can now consider a transformed sequence of observations $\{Y_n\}_{n \geq 2}$, which are not iid and not even independent anymore. Pre- and post-change densities of \mathbf{Y}^n are equal to

$$\begin{aligned} p_i(\mathbf{Y}^n) &= \frac{1}{(2\pi\sigma_i^2)^{n/2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{k=1}^n (Y_k + \theta)^2 \right\} d\theta \\ &= \frac{1}{(2\pi\sigma_i^2)^{(n-1)/2} \sqrt{n}} \exp \left\{ -\frac{1}{2\sigma_i^2} \sum_{k=1}^n (Y_k - \bar{Y}_n)^2 \right\}, \quad i = \infty, 0, \end{aligned} \quad (5.3)$$

where $\bar{Y}_n = n^{-1} \sum_{k=1}^n Y_k$. Define $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$, $s_n^2 = (n-1)^{-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$, and $V_n = (n-1)s_n^2 - (n-2)s_{n-1}^2$. Using (5.3) and noting that $\sum_{k=1}^n (Y_k - \bar{Y}_n)^2 = (n-1)s_n^2$, we obtain

$$p_i(Y_j | \mathbf{Y}^{j-1}) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \sqrt{\frac{j-1}{j}} e^{-V_j/2\sigma_i^2}, \quad j \geq 2,$$

and therefore, the invariant LLR is

$$\lambda_{k+n}^k = \sum_{j=k+1}^{k+n} \log \frac{p_0(Y_j | \mathbf{Y}^{j-1})}{p_\infty(Y_j | \mathbf{Y}^{j-1})} = \frac{q^2 - 1}{2\sigma_0^2} \sum_{j=k+1}^{k+n} V_j - (n-1) \log q, \quad k+n \geq 2 \quad (\lambda_1^k = 0),$$

where $q = \sigma_0/\sigma_\infty$. Taking into account that $\sum_{j=1}^{k+n} V_j = (k+n-1)s_{k+n}^2$, we have

$$\sum_{j=k+1}^{k+n} V_j = (k+n-1)s_{k+n}^2 - ks_{k+1}^2,$$

so the LLR can be written as

$$\lambda_{k+n}^k = \frac{q^2 - 1}{2\sigma_0^2} S_{k,n}^2 - (n-1) \log q, \quad S_{k,n}^2 = (k+n-1)s_{k+n}^2 - ks_{k+1}^2. \quad (5.4)$$

Thus, we can now construct the invariant Shiryaev and SR procedures based on the LLRs λ_n^k , $n \geq 2$ defined in (5.4).

Note first that $S_{k,n}^2/n \rightarrow \sigma_0^2$ as $n \rightarrow \infty$ almost surely under P_k , so that

$$n^{-1} \lambda_{k+n}^k \xrightarrow[n \rightarrow \infty]{P_k\text{-a.s.}} \frac{q^2 - 1}{2} - \log q = I,$$

and I is positive for any $q \neq 1$ ($q > 0$). Thus, condition (3.1) holds with $I = (q^2 - 1)/2 - \log q$ and to apply the results of previous sections it suffices to show that for some $r \geq 1$ and any $\varepsilon > 0$

$$\sum_{n=1}^{\infty} n^{r-1} \sup_{k \in \mathbb{Z}_+} P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right) < \infty. \quad (5.5)$$

To this end, note that the statistic $S_{k,n}^2$ can be written as

$$S_{k,n}^2 = \sum_{i=k+1}^{k+n} (X_i - \bar{X}_{k+n}^{k+1})^2 + k(\bar{X}_{k+n} - \bar{X}_k)^2 + n(\bar{X}_{k+n} - \bar{X}_{k+n}^{k+1})^2,$$

where $\bar{X}_{k+n}^{k+1} = n^{-1} \sum_{i=k+1}^{k+n} X_i$. Denoting

$$k(\bar{X}_{k+n} - \bar{X}_k)^2 + n(\bar{X}_{k+n} - \bar{X}_{k+n}^{k+1})^2 = W_{k,n},$$

and using the fact that $W_{k,n} \geq 0$, we obtain that for some positive $\tilde{\varepsilon}$

$$\begin{aligned} P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right) &= P_k \left(\frac{1}{n} \sum_{i=k+1}^{k+n} (X_i - \bar{X}_{k+n}^{k+1})^2 < \sigma_0^2 - \frac{2\sigma_0^2}{(q^2 - 1)n} \log q - \frac{1}{n} W_{k,n} - \tilde{\varepsilon} \right) \\ &\leq P_k \left(\frac{1}{n} \sum_{i=k+1}^{k+n} (X_i - \bar{X}_{k+n}^{k+1})^2 < \sigma_0^2 - \tilde{\varepsilon} \right) \\ &= P_0 \left(\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 < \sigma_0^2 - \tilde{\varepsilon} \right) \\ &= P_0 \left(\frac{1}{n} (n-1) s_n^2 < \sigma_0^2 - \tilde{\varepsilon} \right). \end{aligned}$$

Since $(n-1)s_n^2/\sigma_0^2$ has chi-squared distribution with $n-1$ degrees of freedom, $P_0((n-1)s_n^2/n - \sigma_0^2 < -\tilde{\varepsilon})$ vanishes exponentially fast as $n \rightarrow \infty$ and it follows that for all $\tilde{\varepsilon} > 0$ and all $r \geq 1$

$$\sum_{n=1}^{\infty} n^{r-1} \sup_{k \in \mathbb{Z}_+} P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right) \leq \sum_{n=1}^{\infty} n^{r-1} P_0 \left(\frac{1}{n} (n-1) s_n^2 < \sigma_0^2 - \tilde{\varepsilon} \right) < \infty.$$

This implies (5.5) for all $r \geq 1$.

By Theorem 3.1 and Theorem 3.2, the Shiryaev detection procedure minimizes asymptotically all positive moments of the detection delay $E^\pi[(T_A - \nu)^m | T_A > \nu]$ (for all $m \geq 1$), and the results of Section 4 for the SR procedure can also be applied to all positive moments of the detection delay.

Example 5.3 (Detection of a change in the correlation coefficient of the AR(1) process). Let the observations represent the Markov Gaussian sequence with the correlation coefficient β_0 before change and β_1 after change, i.e.,

$$X_n = (\beta_0 \mathbb{1}_{\{\nu \leq n\}} + \beta_1 \mathbb{1}_{\{\nu > n\}}) X_{n-1} + w_n, \quad n \geq 1, \quad X_0 = 0,$$

where $w_n \sim \mathcal{N}(0, 1)$, $n \geq 1$ are iid standard normal random variables. Let $|\beta_i| < 1$, $i = 0, 1$, so that the AR(1) process is stable. The pre- and post-change conditional densities are

$$p_\infty(X_n | \mathbf{X}_{n-1}) = \varphi(X_n - \beta_0 X_{n-1}) \quad \text{and} \quad p_0(X_n | \mathbf{X}_{n-1}) = \varphi(X_n - \beta_1 X_{n-1}).$$

The stationary distribution $G(x) = P(X_\infty \leq x)$ of the Markov process $\{X_n\}_{n > k}$ under P_k is given by the random variable $X_\infty = \sum_{n=1}^\infty \beta_1^{n-1} w_n$. Clearly, X_∞ is zero-mean normal with variance $(1 - \beta_1)^{-2}$.

The LLR can be written as

$$\lambda_{k+n}^k = \sum_{i=k+1}^{k+n} g(X_i, X_{i-1}),$$

where

$$g(y, x) = \log \left[\frac{(y - \beta_0 x)^2 - (y - \beta_1 x)^2}{2} \right].$$

Define

$$\tilde{g}(x) = \int_{-\infty}^{\infty} g(y, x) \varphi(y - \beta_1 x) dy = \frac{(\beta_1 - \beta_0)^2 x^2}{2}.$$

We have

$$\sup_{y, x \in (-\infty, \infty)} \frac{|g(y, x)|}{1 + |y|^2 + |x|^2} \leq Q \quad \text{and} \quad \sup_{x \in (-\infty, \infty)} \frac{\tilde{g}(x)}{1 + |x|^2} \leq Q, \quad (5.6)$$

where

$$Q = \max \left\{ 1, \frac{|\beta_1^2 - \beta_0^2| + (\beta_1 - \beta_0)^2 + 1}{2} \right\}.$$

Now, define the Lyapunov function $V(x) = Q(1 + |x|^2)$. Obviously,

$$\lim_{|x| \rightarrow \infty} \frac{E_{x,0} V(X_1)}{V(x)} = \lim_{|x| \rightarrow \infty} \frac{1 + E|\beta_1 x + w_1|^2}{1 + |x|^2} = \beta_1^2 < 1,$$

where $E_{x,0}$ is the expectation under $P_0(\cdot | X_0 = x)$. Therefore, for any $\beta_1^2 < \delta < 1$ there exists $D > 0$ such that condition (C₁) in Section 5 in [Pergamenchchikov and Tartakovsky \(2016\)](#) holds with $C = [-n, n]$ for all $n \geq 1$. Next, by the ergodicity properties, for all $r \geq 1$,

$$\lim_{n \rightarrow \infty} E_{x,0} |X_n|^r = E|X_\infty|^r < \infty \quad \text{for any } x \in (-\infty, \infty), \quad (5.7)$$

where finiteness of $E|X_\infty|^r$ for all $r \geq 1$ follows from the fact that X_∞ is a Gaussian random variable. Observe that under $P_{x,0}$ for any $n \geq 1$

$$X_n = \beta_1^n x + \sum_{i=1}^n \beta_1^{n-i} w_i,$$

and hence, for any $r \geq 1$,

$$\mathbb{E}_{x,0}|X_n|^r \leq 2^r (|x|^r + \mathbb{E}_{0,0}|X_n|^r).$$

Using (5.7), we obtain that for some $C^* > 0$

$$M^*(x) = \sup_{n \geq 1} \mathbb{E}_{x,0}|X_n|^r \leq C^*(1 + |x|^r) \quad \text{and} \quad \sup_{n \geq 1} \mathbb{E}_0 M^*(X_n) < \infty.$$

Therefore, the upper bounds in (5.6) imply condition (C₂) in Section 5 in [Pergamenchchikov and Tartakovsky \(2016\)](#).

By a slight extension of Theorem 5.1 in [Pergamenchchikov and Tartakovsky \(2016\)](#) to $r > 1$,

$$\sum_{n=1}^{\infty} n^{r-1} \sup_{k \in \mathbb{Z}_+} \mathbb{P}_k \left(\left| \frac{1}{n} \lambda_{k+n}^k - I \right| > \varepsilon \right) < \infty \quad \text{for all } r \geq 1 \text{ and } \varepsilon > 0,$$

where

$$I = \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} g(y, x) \varphi(y - \beta_1 x) dy \right) G(dx),$$

i.e., $n^{-1} \lambda_{k+n}^k$ converges r -completely to I as $n \rightarrow \infty$ under \mathbb{P}_k for all $k \in \mathbb{Z}_+$ and all $r \geq 1$ (and even uniformly r -completely). Since the stationary distribution $G(x) = \mathbb{P}(X_{\infty} \leq x)$ of the Markov process X_n under \mathbb{P}_0 is normal $\mathcal{N}(0, (1 - \beta_1)^{-2})$, performing integration we obtain

$$I = \frac{(\beta_1 - \beta_0)^2}{2(1 - \beta_1^2)}.$$

By Theorem 3.1 and Theorem 3.2, the Shiryaev detection procedure minimizes asymptotically $\mathbb{E}^{\pi}[(T - \nu)^m | T > \nu]$ for all $m \geq 1$, and the results of Section 4 for the SR procedure can also be applied to all positive moments of the detection delay.

6. Concluding remarks and discussion

1. The performed study shows that the famous Shiryaev and Shiryaev–Roberts change detection procedures have certain interesting asymptotic properties in the Bayesian context. Specifically, the Shiryaev procedure is asymptotically optimal (when the probability of false alarm is small) with respect to moments of the detection delay up to order $r \geq 1$ for general non-iid models under mild conditions. These conditions are expressed via the SLLN for the LLR process and a rate of convergence (r -complete convergence). The r -complete convergence is usually not difficult to check in particular applications and examples. On the other hand, the r -quick convergence condition previously used in [Tartakovsky and Veeravalli \(2005\)](#) is stronger, and more importantly, usually much more difficult to verify.

2. A detailed examination of the proofs in Section 3.1 shows that the Shiryaev procedure $T_{A_{\alpha}}$ minimizes not only the “average” moments $\mathbb{E}^{\pi}[(T - \nu)^r | T > \nu]$ but also conditional moments $\mathbb{E}_{\nu}[(T - \nu)^r | T > \nu]$ uniformly for all (fixed) change points $\nu = 0, 1, 2, \dots$ in class \mathbb{C}_{α} asymptotically as $\alpha \rightarrow 0$. Specifically, with an additional effort it can be established that under the conditions of Theorem 3.1 for all $\nu \in \mathbb{Z}_+$ as $\alpha \rightarrow 0$

$$\inf_{T \in \mathbb{C}_{\alpha}} \mathbb{E}_{\nu}[(T - \nu)^r | T > \nu] \sim \mathbb{E}_{\nu}[(T_{A_{\alpha}} - \nu)^r | T_{A_{\alpha}} > \nu] \sim \left(\frac{|\log \alpha|}{I + \mu} \right)^r.$$

3. The study of the Shiryaev–Roberts procedure shows that it is suboptimal in the Bayesian problem if the prior distribution has an exponential tail, but remains asymptotically optimal when the tail is heavy or if the parameter that characterizes the exponential tail goes to zero. This is expected since the SR procedure does not use the given prior distribution.

4. [Lai \(1998\)](#) proved asymptotic optimality of the CUSUM procedure with respect to the expected detection delay $\sum_{k=0}^{\infty} \pi_k^{\alpha} \mathbb{E}_k(T - k)^+$ in class \mathbb{C}_{α} under the following essential supremum condition

$$\lim_{n \rightarrow \infty} \sup_{\ell \geq k} \text{ess sup } \mathbb{P}_k \left(\lambda_{\ell+n}^{\ell} - I \leq -\varepsilon | \mathcal{F}_{\ell} \right) = 0 \quad \text{for all } \varepsilon > 0.$$

However, on one hand, this condition is much more difficult to verify than the complete convergence condition required in our theorems. On the other hand, for many interesting models (including Markov and hidden Markov models) Lai's condition does not hold, while the complete convergence condition holds. This is the case, e.g., in [Example 5.3](#).

Appendix: Proofs of Lemmas

Proof of Lemma 3.1. For $\varepsilon \in (0, 1)$, define $N_{\alpha, \varepsilon} = (1 - \varepsilon)|\log \alpha|/(I + \delta)$. By the Chebyshev inequality,

$$\begin{aligned} \mathbb{E}^{\pi}[(T - \nu)^m | T > \nu] &\geq \mathbb{E}^{\pi}[(T - \nu)^+]^m \geq N_{\alpha, \varepsilon}^m \mathbb{P}^{\pi}(T - \nu > N_{\alpha, \varepsilon}) \\ &\geq N_{\alpha, \varepsilon}^m [\mathbb{P}^{\pi}(T > \nu) - \mathbb{P}^{\pi}(0 < T - \nu < N_{\alpha, \varepsilon})] \end{aligned}$$

where

$$\mathbb{P}^{\pi}(0 < T - \nu < N_{\alpha, \varepsilon}) = \sum_{k=0}^{\infty} \pi_k \mathbb{P}_k(k < T < k + N_{\alpha, \varepsilon}).$$

Since for any $T \in \mathbb{C}_{\alpha}$, $\mathbb{P}^{\pi}(T > \nu) = 1 - \text{PFA}(T) \geq 1 - \alpha$, we obtain

$$\inf_{T \in \mathbb{C}_{\alpha}} \mathbb{E}^{\pi}[(T - \nu)^m | T > \nu] \geq N_{\alpha, \varepsilon}^m \left[1 - \alpha - \sup_{T \in \mathbb{C}_{\alpha}} \mathbb{P}^{\pi}(0 < T - \nu < N_{\alpha, \varepsilon}) \right]. \quad (\text{A.1})$$

Thus, to prove the lower bound (3.2) we need to show that

$$\lim_{\alpha \rightarrow 0} \sup_{T \in \mathbb{C}_{\alpha}} \mathbb{P}^{\pi}(0 < T - \nu < N_{\alpha, \varepsilon}) = 0. \quad (\text{A.2})$$

To this end, introduce

$$U_{\alpha, \varepsilon}^k(T) = e^{(1+\varepsilon)IN_{\alpha, \varepsilon}} \mathbb{P}_{\infty}(k < T < k + N_{\alpha, \varepsilon}), \quad \beta_{\alpha, \varepsilon}^k = \mathbb{P}_k \left(\frac{1}{N_{\alpha, \varepsilon}} \max_{1 \leq n \leq N_{\alpha, \varepsilon}} \lambda_{k+n}^k \geq (1 + \varepsilon)I \right).$$

By inequality (3.6) in [Tartakovsky and Veeravalli \(2005\)](#),

$$\mathbb{P}_k(k < T < k + N_{\alpha, \varepsilon}) \leq U_{\alpha, \varepsilon}^k(T) + \beta_{\alpha, \varepsilon}^k. \quad (\text{A.3})$$

It is easy to see that

$$\sup_{T \in \mathbb{C}_{\alpha}} \mathbb{P}_{\infty}(T \leq k) \leq \alpha / \mathbb{P}(\nu \geq k), \quad k \geq 1$$

(cf. (3.8) in [Tartakovsky and Veeravalli \(2005\)](#)), so that

$$\begin{aligned} U_{\alpha, \varepsilon}^k(T) &\leq e^{(1+\varepsilon)IN_{\alpha, \varepsilon}} \mathbb{P}_{\infty}(T < k + N_{\alpha, \varepsilon}) \leq \alpha e^{(1+\varepsilon)IN_{\alpha, \varepsilon}} / \mathbb{P}(\nu > k + N_{\alpha, \varepsilon}) \\ &\leq \exp \left\{ (1 + \varepsilon)IN_{\alpha, \varepsilon} - |\log \alpha| - (k + N_{\alpha, \varepsilon}) \frac{\log \mathbb{P}(\nu > k + N_{\alpha, \varepsilon})}{k + N_{\alpha, \varepsilon}} \right\}. \end{aligned}$$

By condition (2.3), for all sufficiently large $N_{\alpha, \varepsilon}$ (small α), there exists a (small) δ such that

$$-\frac{\log \mathbb{P}(\nu > k + N_{\alpha, \varepsilon})}{k + N_{\alpha, \varepsilon}} \leq \mu + \delta.$$

Hence, for a sufficiently small α ,

$$\begin{aligned} \sup_{T \in \mathbb{C}_\alpha} U_{\alpha, \varepsilon}^k(T) &\leq \exp \{ (1 + \varepsilon) I N_{\alpha, \varepsilon} - |\log \alpha| + (k + N_{\alpha, \varepsilon})(\mu + \delta) \} \\ &\leq \exp \{ -\varepsilon^2 |\log \alpha| + (\mu + \delta)(k + N_{\alpha, \varepsilon}) \}, \end{aligned}$$

which approaches zero as $\alpha \rightarrow 0$ for $k \leq K_{\alpha, \varepsilon} = \varepsilon_1 \varepsilon^2 |\log \alpha| / (\mu + \delta)$, where $0 < \varepsilon_1 < 1$ and $\delta \rightarrow 0$ as $\alpha \rightarrow 0$. By condition (3.1), $\beta_{\alpha, \varepsilon}^k \rightarrow 0$ for all $k \in \mathbb{Z}_+$, and therefore, we obtain

$$\begin{aligned} \sup_{T \in \mathbb{C}_\alpha} \mathbb{P}^\pi(0 < T - \nu < N_{\alpha, \varepsilon}) &= \sum_{k=0}^{\infty} \pi_k \sup_{T \in \mathbb{C}_\alpha} \mathbb{P}_k(k < T < k + N_{\alpha, \varepsilon}) \\ &\leq \mathbb{P}(\nu > K_{\alpha, \varepsilon}) + \sum_{k=0}^{K_{\alpha, \varepsilon}} \pi_k \beta_{\alpha, \varepsilon}^k + \max_{0 \leq k \leq K_{\alpha, \varepsilon}} \sup_{T \in \mathbb{C}_\alpha} U_{\alpha, \varepsilon}^k(T) \\ &\leq \mathbb{P}(\nu > K_{\alpha, \varepsilon}) + \sum_{k=0}^{K_{\alpha, \varepsilon}} \pi_k \beta_{\alpha, \varepsilon}^k + \exp \{ -\varepsilon^2 |\log \alpha| + (\mu + \delta) K_{\alpha, \varepsilon} \}, \quad (\text{A.4}) \end{aligned}$$

where all three terms go to zero as $\alpha \rightarrow 0$ for all $\varepsilon > 0$, so that (A.2) follows and the proof of the lower bound (3.2) is complete.

The proof of the lower bound (3.3) is essentially similar. Indeed, recall that, by (2.10), $T_A \in \mathbb{C}_\alpha$ if $A = 1/\alpha$, so that it suffices to replace α by $1/A$ in the above argument. The details are omitted. \square

Proof of Lemma 3.2. Obviously, for any $n > k$,

$$\log \Lambda_n \geq \log \left(\frac{\pi_k}{\mathbb{P}^\pi(\nu \geq n)} e^{\lambda_n^k} \right) = \lambda_n^k + \log \pi_k - \log \mathbb{P}(\nu \geq n),$$

and hence, for every $A > 0$,

$$(T_A - k)^+ \leq \tau_A^{(k)} := \inf \left\{ n \geq 1 : \lambda_{k+n}^k + |\log \mathbb{P}(\nu \geq k + n)| \geq \log(A/\pi_k) \right\}, \quad k \in \mathbb{Z}_+,$$

and $\mathbb{E}_k[(T_A - k)^+]^r \leq \mathbb{E}_k(\tau_A^{(k)})^r$.

Let $N_A = 1 + \lfloor \log(A/\pi_k)/(I + \mu - \varepsilon) \rfloor$, where $\lfloor x \rfloor$ is the greatest integer less or equal to x . For any $k \in \mathbb{Z}_+$, we have

$$\begin{aligned} \mathbb{E}_k[(T_A - k)^+]^r &\leq \mathbb{E}_k \left(\tau_A^{(k)} \right)^r = \int_0^\infty r t^{r-1} \mathbb{P}_k \left(\tau_A^{(k)} > t \right) dt \\ &\leq N_A^r + \sum_{n=0}^\infty \int_{N_A+n}^{N_A+n+1} r t^{r-1} \mathbb{P}_k \left(\tau_A^{(k)} > t \right) dt \\ &\leq N_A^r + \sum_{n=0}^\infty \int_{N_A+n}^{N_A+n+1} r t^{r-1} \mathbb{P}_k \left(\tau_A^{(k)} > N_A + n \right) dt \\ &= N_A^r + \sum_{n=0}^\infty [(N_A + n + 1)^r - (N_A + n)^r] \mathbb{P}_k \left(\tau_A^{(k)} > N_A + n \right) \\ &= N_A^r + \sum_{n=N_A}^\infty [(n+1)^r - n^r] \mathbb{P}_k \left(\tau_A^{(k)} > n \right) \\ &\leq N_A^r + \sum_{n=N_A}^\infty r(n+1)^{r-1} \mathbb{P}_k \left(\tau_A^{(k)} > n \right) \end{aligned}$$

$$\leq N_A^r + \sum_{n=N_A}^{\infty} r 2^{r-1} n^{r-1} P_k \left(\tau_A^{(k)} > n \right). \quad (\text{A.5})$$

It is easily seen that for all $k \in \mathbb{Z}_+$ and $n \geq N_A$

$$\begin{aligned} P_k \left(\tau_A^{(k)} > n \right) &\leq P_k \left(\frac{1}{n} \lambda_{k+n}^k < \frac{1}{n} \log(A/\pi_k) - \frac{1}{n} |\log P(\nu \geq k+n)| \right) \\ &\leq P_k \left(\frac{1}{n} \lambda_{k+n}^k < I + \mu - \varepsilon - \frac{1}{n} |\log P(\nu \geq k+n)| \right). \end{aligned} \quad (\text{A.6})$$

Since by condition (C), $N_A^{-1} |\log P(\nu \geq k+N_A)| \rightarrow \mu$ as $A \rightarrow \infty$, for a sufficiently large value of A there exists a small $\delta = \delta_A$ ($\delta_A \rightarrow 0$ as $A \rightarrow \infty$) such that $|\mu - |\log P(\nu \geq k+N_A)|/N_A| < \delta$. Hence, for all sufficiently large A ,

$$P_k \left(\tau_A^{(k)} > n \right) \leq P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon - \delta \right). \quad (\text{A.7})$$

Using (A.5) and (A.7), we obtain

$$\begin{aligned} E_k \left(\tau_A^{(k)} \right)^r &\leq N_A^r + r 2^{r-1} \sum_{n=N_A}^{\infty} n^{r-1} P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon - \delta \right) \\ &\leq N_A^r + r 2^{r-1} \sum_{n=1}^{\infty} n^{r-1} P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon - \delta \right). \end{aligned}$$

This implies inequality (3.9) in Lemma 3.2.

If $\pi_k = \rho(1-\rho)^k$ is geometric, condition (2.3) holds for all $n \geq 1$ with $\mu = |\log(1-\rho)|$, so that

$$\log(A/\pi_k) = \log(A/\rho) + \mu k \quad \text{and} \quad |\log P(\nu \geq k+n)| = -\log(1-\rho)^{k+n} = (k+n)\mu.$$

Therefore, the Markov times $\tau_A^{(k)}$, $k = 0, 1, \dots$ can be written as

$$\tau_A^{(k)} = \inf \left\{ n \geq 1 : \lambda_{k+n}^k + \mu n \geq \log(A/\rho) \right\},$$

and inequality (A.6) reduces to

$$P_k \left(\tau_A^{(k)} > n \right) \leq P_k \left(\frac{1}{n} \lambda_{k+n}^k < I - \varepsilon \right),$$

which holds for all $n \geq N_A = 1 + \lfloor \log(A/\rho)/(I + \mu - \varepsilon) \rfloor$ and all $0 < \varepsilon < I + \mu$. Using this inequality and inequality (A.5) with $N_A = 1 + \lfloor \log(A/\rho)/(I + \mu - \varepsilon) \rfloor$ yields inequality (3.10) in Lemma 3.2 and the proof is complete. \square

Proof of Lemma 3.3. Let $N_{\alpha, \varepsilon} = (1-\varepsilon)|\log \alpha|/(I + \mu + \delta)$. The rest of the notation is the same as in the proof of Lemma 3.1 above. Similarly to (A.1),

$$\inf_{T \in \mathbb{C}_\alpha} E^{\pi^\alpha} [(T - \nu)^m | T > \nu] \geq N_{\alpha, \varepsilon}^m \left[1 - \alpha - \sup_{T \in \mathbb{C}_\alpha} P^{\pi^\alpha} (0 < T - \nu < N_{\alpha, \varepsilon}) \right]. \quad (\text{A.8})$$

By (A.4),

$$\sup_{T \in \mathbb{C}_\alpha} P^{\pi^\alpha} (0 < T - \nu < N_{\alpha, \varepsilon}) \leq P(\nu > K_{\alpha, \varepsilon}) + \sum_{k=0}^{K_{\alpha, \varepsilon}} \pi_k^\alpha \beta_{\alpha, \varepsilon}^k + \exp \{ -\varepsilon^2 |\log \alpha| + (\mu_\alpha + \delta_\alpha) K_{\alpha, \varepsilon} \},$$

where $K_{\alpha,\varepsilon} = \varepsilon_1 \varepsilon^2 |\log \alpha| / (\mu_\alpha + \delta_\alpha)$ with $0 < \varepsilon_1 < 1$ (in particular, we may take $\varepsilon_1 = \varepsilon$). Obviously, the last term vanishes as $\alpha \rightarrow 0$ for all $\varepsilon \in (0, 1)$. By condition (3.1), the middle term also goes to zero. By condition (2.3) on the prior, as $\alpha \rightarrow 0$,

$$-\log P(\nu > K_{\alpha,\varepsilon}) \sim \mu_\alpha K_{\alpha,\varepsilon} \sim \varepsilon_1 \varepsilon^2 |\log \alpha| \rightarrow \infty.$$

Therefore, the first term $P(\nu > K_{\alpha,\varepsilon})$ also approaches zero as $\alpha \rightarrow 0$ for all $\varepsilon \in (0, 1)$. It follows that

$$\sup_{T \in \mathbb{C}_\alpha} P^{\pi^\alpha}(0 < T - \nu < N_{\alpha,\varepsilon}) \rightarrow 0 \quad \text{as } \alpha \rightarrow 0$$

and using (A.8), we obtain that for all $0 < \varepsilon < 1$ and $m > 0$ as $\alpha \rightarrow 0$

$$\inf_{T \in \mathbb{C}_\alpha} E^{\pi^\alpha}[(T - \nu)^m | T > \nu] \geq (1 - \varepsilon)^m \left(\frac{|\log \alpha|}{I} \right)^m (1 + o(1)).$$

Since ε can be arbitrarily small, the lower bound (3.15) follows.

To prove the lower bound (3.16) it suffices to replace α in the above argument by $1/A$ and recall that $T_A \in \mathbb{C}_\alpha$ if $A = 1/\alpha$. \square

References

- Baron, M. and Tartakovsky, A.G. (2006). Asymptotic Bayesian change-point detection theory for general continuous-time models. *Sequential Analysis*, **25**, 257–296.
- Basseville, M. and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Applications*. Englewood Cliffs: Prentice Hall.
- Fuh, C.D. (2003). SPRT and CUSUM in hidden Markov models. *Ann. Statist.*, **31**, 942–977.
- Girshick, M.A. and Rubin, H. (1952). A Bayes approach to a quality control model. *Ann. Math. Statist.*, **23**, 114–125.
- Hsu, P.L. and Robbins, H. (1947). Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. U.S.A.*, **33**, 25–31.
- Lai, T.L. (1976). On r -quick convergence and a conjecture of Strassen. *Ann. Probab.*, **4**, 612–627.
- Lai, T.L. (1998). Information bounds and quick detection of parameter changes in stochastic systems. *IEEE Trans. Inform. Theory*, **44**, 2917–2929.
- Moustakides, G.V., Polunchenko, A.S., and Tartakovsky, A.G. (2011). A numerical approach to performance analysis of quickest change-point detection procedures. *Statist. Sinica*, **21**, 571–596.
- Pergamenchchikov, S. and Tartakovsky, A.G. (2016). Asymptotically optimal pointwise and minimax quickest change-point detection for dependent data. *Statist. Infer. Stoch. Proc.*, Submitted.
- Pollak, M. and Tartakovsky, A.G. (2009). Optimality properties of the Shiryaev–Roberts procedure. *Statist. Sinica*, **19**, 1729–1739.
- Polunchenko, A.S. and Tartakovsky, A.G. (2010). On optimality of the Shiryaev–Roberts procedure for detecting a change in distribution. *Ann. Statist.*, **38**, 3445–3457.
- Roberts, S.W. (1966). A comparison of some control chart procedures, *Technometrics*, **8**, 411–430.
- Shiryaev, A.N. (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Dokl. Math.*, **2**, 795–799.
- Shiryaev, A.N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.*, **8**, 22–46.
- Tartakovsky, A.G. and Moustakides, G.V. (2010). State-of-the-art in Bayesian changepoint detection. *Sequential Analysis*, **29**, 125–145.
- Tartakovsky, A., Nikiforov, I., and Basseville, M. (2014). *Sequential Analysis: Hypothesis Testing and Changepoint Detection*. (Monographs on Statistics and Applied Probability), Boca Raton, London, New York: Chapman & Hall/CRC Press.
- Tartakovsky, A.G. and Veeravalli, V.V. (2005). General asymptotic Bayesian theory of quickest change detection. *Theory Probab. Appl.*, **49**, 458–497.
- Tartakovsky, A.G., Pollak, M. and Polunchenko, A.S. (2011). Third-order asymptotic optimality of the generalized Shiryaev–Roberts changepoint detection procedures. *Theory Probab. Appl.*, **56**, 534–565.